Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback

Sebastian Wagner, *Student Member, IEEE,* Romain Couillet, *Student Member, IEEE,* Mérouane Debbah, *Senior Member, IEEE,* and Dirk T. M. Slock, *Fellow, IEEE*

***Abstract*—In this paper we study the sum rate of zero-forcing (ZF) as well as regularized ZF (RZF) precoding in large MISO broadcast channels, under the assumptions of imperfect channel state information at the transmitter, channel transmit correlation and different user path losses. Our analysis assumes that the number of transmit antennas M and the number of users K are large and of the same order of magnitude. We apply recent results on the empirical spectral distribution of certain kinds of large dimensional random matrices to derive deterministic equivalents for the signal-to-interference plus noise ratio (SINR) at the receivers. Based on these results and under sum rate maximization, we evaluate for RZF (i) the optimal precoder, for ZF (ii) the optimal number of active users and (iii) the optimal amount of channel training in TDD multi-user systems. Moreover, we study the achievable sum rate under limited feedback and derive an approximation of the feedback rate required to maintain a given rate offset relative to perfect CSIT. Numerical simulations suggest that the approximations, almost surely exact as $M,K\to\infty$, are accurate even for small M,K.**

***Index Terms*—Broadcast channel, random matrix theory, linear precoding, limited feedback, multi-user.**

I. Introduction

The pioneering work in [1] and [2] revealed that the capacity of a point-to-point (single-user (SU)) multiple-input multiple-output (MIMO) channel can potentially increase linearly with the number of antennas. However, practical implementations quickly demonstrated that in most propagation environments the promised capacity gain of SU-MIMO is unachievable due to antenna correlation and line-of-sight components [3]. In a multi-user scenario, the inherent problems of SU-MIMO transmission can largely be overcome by exploiting multi-user (MU) diversity, i.e. sharing the spatial dimension not only between the antennas of a single receiver, but among multiple (non-cooperative) users. The underlying channel for MU-MIMO transmission is referred to as the MIMO broadcast channel (BC) or MU downlink channel. Although much more robust to channel correlation, the MIMO-BC suffers from inter-user interference at the receivers which can only be efficiently mitigated by appropriate (i.e. channel-aware) pre-processing at the transmitter.

It has been proved that dirty-paper coding (DPC) is a capacity achieving precoding strategy for the Gaussian MIMO-BC [4]–[8]. But the DPC precoder is non-linear and to this day too complex to be implemented efficiently in practical systems. However, it has been shown in [4], [9]–[11], that suboptimal *linear* precoders can achieve a large portion of the BC rate region while featuring low computational complexity. Thus, a lot of research has recently focused on linear precoding strategies.

In general, the rate maximizing linear precoder has no explicit form. Several iterative algorithms have been proposed in [12], [13], but no global convergence has been proved. Still, these iterative algorithms have a high computational complexity which motivates the use of further suboptimal linear transmit filters (i.e. precoders), by imposing more structure into the filter design. A straightforward technique is to precode by the inverse of the channel matrix. This scheme is usually referred to as channel inversion or zero-forcing (ZF) [4]. Similar to the approach pursued in the present contribution, the authors in [14], [15] carry out a large system analysis assuming that the number of transmit antennas M as well as the number of users K grow large while their ratio $\beta\triangleq M/K$ remains bounded. It is shown in [14] that for $\lim_M \beta > 1$, ZF achieves a large fraction of the linear (w.r.t. K) sum rate growth. The work in [9] extends the analysis in [14] to the case $\beta=1$ and shows that the sum rate of ZF is constant in K as $K\to\infty$; the linear sum rate growth is lost due to the large ratio of the maximum to minimum singular value of the channel matrix. The authors in [9] counter this problem by introducing a regularization term into the inverse of the channel matrix. Under the assumption of large K and for any rotationally-invariant channel distribution, [9] derives the regularization term that maximizes the signal-to-interference plus noise ratio (SINR). In this article, the resulting regularized ZF (RZF) precoders are referred to as *channel distortion-unaware* RZF (RZF-CDU), since their design assumes perfect channel state information at the transmitter (CSIT). It has been observed that the RZF-CDU is very similar to the transmit filter derived under the minimum mean square error (MMSE) criterion [16] and both become identical in the large K limit. Likewise, we will observe some similarities between RZF and MMSE filters when considering imperfect channel state information at the transmitter (CSIT).

In the large system limit, i.e. $K,M\to\infty$, and for independent and identically distributed (i.i.d.) channels the cross correlations between the user channels, and therefore the user SINRs, are identical. It has been shown in [17] that for this symmetric case and equal noise variances, the SINR maximizing precoder is of closed form and coincides with RZF precoder. This asymptotic optimality motivates a detailed

S. Wagner and R. Couillet are with ST-ERICSSON, Sophia-Antipolis

S. Wagner and D.T.M. Slock are with EURECOM, Sophia-Antipolis, 06904, Route des Crêtes, B.P. 193, France e-mail:{sebastian.wagner, dirk.slock}@eurecom.fr.

R. Couillet and M. Debbah are with the Alcatel-Lucent Chair on Flexible Radio, SUPÉLEC, Gif sur Yvette, 91192, Plateau de Moulon, 3, Rue Joliot-Curie, France e-mail:{romain.couillet, merouane.debbah}@supelec.fr.

analysis of the RZF precoder for large system dimensions. The RZF-CDU has been analyzed for i.i.d. channels in a large system in [18] for a single-cell setup and in [19] for a two-cell system. In [18] the authors derive the optimal regularization parameter using an asymptotic expression for the SINR. The expressions in [18] are a special case of the results derived in this paper. However, the approach and the tools for the derivations of the asymptotic SINR in [18] have been independently introduced earlier in the context of a different RZF precoder [20], where the regularization parameter is set to fulfill the total transmit power constraint.

Although it is legitimate that [9], [12], [13], [16] assume perfect CSIT to determine theoretically optimal performance, this assumption is untenable in practice. Also, it is a particularly strong assumption, since the performance of all precoding strategies is crucially depending on the CSIT. In practical systems, the transmitter has to acquire the channel state information (CSI) of the downlink channel by feedback signaling from the uplink. Since in practice the channel coherence time is finite, the information of the instantaneous channel state is inherently incomplete. For this reason, a lot of research has been carried out to understand the impact of imperfect CSIT on the system behavior, cf. [21] for a recent survey.

An information theoretic analysis of the impact of imperfect CSIT on the achievable rate of a ZF precoded MU-MISO downlink channel with $\beta = 1$ has been carried out in [22]. Hereby, the author derives an upper bound on the per-user rate gap between perfect CSIT and imperfect CSIT under random vector quantization (RVQ) with B feedback bits per user. Under finite-rate feedback, both [22] and [23] observe a sum-rate ceiling for high signal-to-noise ratios (SNR). In [22, Theorem 3] provides a formula for the minimum scaling of B to maintain a per-user rate gap of $\log_2 b$ bits/s/Hz. Although derived for ZF, the author claims that for all SNR, [22, Theorem 3] is even more accurate for the RZF-CDU proposed in [9].

This work extends the results in [9], [14], [18], [22], [23] by applying novel results of large dimensional random matrices to derive deterministic approximations of the SINR under ZF and RZF precoding. These approximations are referred to as *deterministic equivalents* as they are *independent* of the stochastic parameters of the system model. Moreover, these deterministic equivalents match the true SINR, almost surely, for asymptotically large K. Also, as corroborated by numerical results in Section VII-B, they approximate the true SINR very accurately even for small K. The communication channel correlation is modeled as Kronecker to account for transmit correlation and different user path losses. In this general case, these deterministic equivalents do not have a closed form expression but are the solution of an implicit equation. For uncorrelated channels and equal path losses though, the deterministic equivalents have a closed form expression.

This framework directly extends the analysis for ZF in [14] and [18] to include transmit correlation, path losses and imperfect CSIT. Furthermore, our approach allows for a unification and extension of the RZF analysis in [9], [18], [22]. In particular, we optimize the RZF-CDU proposed in [9], [18], where the optimal regularization term is the solution to an implicit equation and has a closed form for uncorrelated signals and equal path losses. This optimal RZF precoder is referred to as ORZF. Moreover, for RZF and ZF, to maintain a per-user rate offset of $\log_2 b$ bits/s/Hz, we derive the required scaling laws of the distortion of the CSIT as a function of the SNR. Under RVQ, these results extend [22, Theorem 3].

Our main contributions can be summarized as follows:

- We derive deterministic equivalents for the SINR of ZF ($M > K$) and RZF ($M \geq K$) precoders, i.e. deterministic approximations of the SINR, which are independent of the individual channel realizations, and asymptotically (almost surely) exact as $M, K \to \infty$. Numerical results prove that these approximations are accurate even for finite (K, M).
- The sum rate of the ORZF saturates under limited feedback at asymptotically high SNR, and we determine the sum rate saturation value.
- Under RVQ, for $\beta = 1$ and high SNR ρ, to maintain a per-user rate offset of $\log_2 b$ bits/s/Hz, the number of feedback bits B per user has to scale approximately with
  - ORZF: $B = (M-1)\log_2 \rho - (M-1)\log_2(b^2-1)$
  - RZF-CDU: $B = (M-1)\log_2 \rho - (M-1)\log_2 2(b-1)$

  That is, the ORZF requires $(M-1)\log \frac{b+1}{2}$ bits *less* than the former RZF-CDU and $(M-1)\log(b+1)$ bits *less* than ZF.

The remainder of the paper is organized as follows. Section II introduces tools from random matrix theory essential to our subsequent derivations. Section III presents the communication channel model. In Section IV, we derive deterministic equivalents for the sum rate of RZF and ZF. In Section V, we analyze the sum rate under limited feedback. Section VI studies several practical applications and discusses the aftermath of our derivations. Section VII presents numerical results comparing the theoretical findings to Monte-Carlo simulations. Finally, in Section VIII, we summarize our results and conclude the paper. Note that Appendix IV collects a set of lemmas and a corollary that will be used throughout the paper.

Notation: In the following boldface lower-case and upper-case characters denote vectors and matrices, respectively. The operators $(\cdot)^{\mathsf{H}}$, $\mathrm{tr}(\cdot)$ and, for $\mathbf{X}$ of size $N \times N$, $\mathrm{Tr}\mathbf{X} \triangleq \frac{1}{N}\mathrm{tr}\mathbf{X}$ and $\|\mathbf{X}\|$ denote conjugate transpose, trace, normalized matrix trace and spectral norm, respectively. The expectation is $E[\cdot]$ and $\mathrm{diag}(x_1, \ldots, x_N)$ is the diagonal matrix with ith diagonal entry x_i. The $N \times N$ identity matrix is $\mathbf{I}_N$ and $\Im[z]$ is the imaginary part of $z \in \mathbb{C}$.

II. Mathematical Preliminaries

In the present work we are interested in deterministic equivalents of functionals of matrices of the form

$$m_{\mathbf{B}_N,\mathbf{Q}_N}(z) \triangleq \mathrm{Tr}\mathbf{Q}_N \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}, \tag{1}$$

where $\mathbf{Q}_N \in \mathbb{C}^{N\times N}$ is a Hermitian positive definite matrix and $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ is of the type

$$\mathbf{B}_N = \mathbf{R}_N^{1/2}\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N\mathbf{R}_N^{1/2} + \mathbf{S}_N, \tag{2}$$

where $\mathbf{R}_N, \mathbf{S}_N \in \mathbb{C}^{N\times N}$ are nonnegative definite Hermitian matrices, $\mathbf{R}_N^{1/2}$ is a Hermitian square-root of $\mathbf{R}_N$, $\mathbf{T}_N \in \mathbb{C}^{n\times n}$ is a nonnegative definite *diagonal* matrix and $\mathbf{X}_N \in \mathbb{C}^{n\times N}$ is random with i.i.d. entries of zero mean and variance $1/N$. In the course of the derivations, we will require the following result,

Theorem 1: Let $\mathbf{B}_N$ be defined as in (2), where we assume that $\mathbf{T}_N$, $\mathbf{R}_N$ and $\mathbf{Q}_N$ have uniformly bounded spectral norm (with respect to N), as (n, N) grow large with ratio $\beta(N) \triangleq N/n$ such that $0 < \liminf_N \beta(N) \leq \limsup_N \beta(N) < \infty$. Define $m_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ as in (1). Then, for $z \in \mathbb{C}\setminus\mathbb{R}^+$,

$$m_{\mathbf{B}_N,\mathbf{Q}_N}(z) - m^\circ_{\mathbf{B}_N,\mathbf{Q}_N}(z) \overset{N\to\infty}{\longrightarrow} 0, \tag{3}$$

almost surely, with $m^\circ_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ given by

$$m^\circ_{\mathbf{B}_N,\mathbf{Q}_N}(z) = \mathrm{Tr}\mathbf{Q}_N \left(c(z)\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1}, \tag{4}$$

and $e(z)$ is the unique solution of

$$e(z) = \mathrm{Tr}\mathbf{R}_N \left(c(z)\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1}, \tag{5}$$

$$c(z) = \frac{1}{\beta}\mathrm{Tr}\mathbf{T}_N \left(\mathbf{I}_n + e(z)\mathbf{T}_N\right)^{-1} \tag{6}$$

with positive imaginary part if $\Im[z] > 0$, with negative imaginary part if $\Im[z] < 0$, or real positive if $z < 0$. Moreover $e(z)$ is analytic on $\mathbb{C}\setminus\mathbb{R}^+$ and of uniformly bounded module on every compact subset of $\mathbb{C}\setminus\mathbb{R}^+$.

The proof of Theorem 1 is provided in Appendix I.

We denote $m_{\mathbf{B}_N,\mathbf{I}_N}(z) \triangleq m_{\mathbf{B}_N}(z)$ the Stieltjes transform [24] of the eigenvalue distribution of $\mathbf{B}_N$, used in random matrix theory, e.g. by Marčenko and Pastur [25] to derive the limiting distribution of sample covariance matrices (which corresponds here to the case $\mathbf{Q}_N = \mathbf{R}_N = \mathbf{S}_N = \mathbf{I}_N$).

Remark 1: If $\mathbf{X}_N$ is Gaussian, besides (3), the authors infer that, by applying the *Gaussian method* [26], it can be shown that

$$K\left[m_{\mathbf{B}_N,\mathbf{Q}_N}(z) - m^\circ_{\mathbf{B}_N,\mathbf{Q}_N}(z)\right] \overset{N\to\infty}{\longrightarrow} 0, \tag{7}$$

almost surely. This result is outside the scope of this paper and will not be proved.

For practical purposes, we prove hereafter that the implicit equation (5) can be solved numerically, when $z < 0$; the parameter z will in particular be linked to the regularization factor for RZF precoders. The algorithm known as the *fixed-point algorithm*, when properly initialized, is shown to converge surely to the unique positive solution of (5).

Proposition 1: Let $z < 0$. Define the sequence $e_0, e_1, \ldots$ as $0 < e_0 \leq -1/z$ and, for $k \geq 0$,

$$e_{k+1} = \mathrm{Tr}\mathbf{R}_N \left(c_k\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N\right)^{-1}, \tag{8}$$

$$c_k = \frac{1}{\beta}\mathrm{Tr}\mathbf{T}_N \left(\mathbf{I}_n + e_k\mathbf{T}_N\right)^{-1}. \tag{9}$$

Then the sequence $e_0, e_1, \ldots$ converges surely to $e(z)$, the unique solution to (5) in Theorem 1.

The proof of Proposition 1 is provided in Appendix II.

III. System Model

This section describes the transmission model as well as the underlying channel model.

A. Transmission Model

Consider the MISO broadcast channel composed of a central transmitter equipped with M antennas and of K single-antenna receivers. Assume narrow-band communication. Denoting y_k the signal received by user k, the concatenated received signal vector $\mathbf{y} = [y_1, \ldots, y_K]^\mathsf{T} \in \mathbb{C}^K$ at a given time instant reads

$$\mathbf{y} = \sqrt{M}\mathbf{H}\mathbf{G}\mathbf{s} + \mathbf{n} \tag{10}$$

with symbol vector $\mathbf{s} = [s_1, \ldots, s_K]^\mathsf{T} \sim \mathcal{CN}(0, \mathbf{I}_K)$, precoding matrix $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M\times K}$, channel matrix $\mathbf{H} \in \mathbb{C}^{K\times M}$ and noise vector $\mathbf{n} = [n_1, \ldots, n_K]^\mathsf{T} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_K)$. Assuming a total transmit power $P > 0$,

$$\mathrm{tr}(E[\mathbf{G}\mathbf{s}\mathbf{s}^\mathsf{H}\mathbf{G}^\mathsf{H}]) = \mathrm{tr}(\mathbf{G}\mathbf{G}^\mathsf{H}) \leq P. \tag{11}$$

We define the SNR ρ as $\rho \triangleq P/\sigma^2$. The received symbol y_k of user k is given by

$$y_k = \sqrt{M}\mathbf{h}_k^\mathsf{H}\mathbf{g}_k s_k + \sqrt{M}\sum_{i=1, i\neq k}^{K} \mathbf{h}_k^\mathsf{H}\mathbf{g}_i s_i + n_k, \tag{12}$$

where $\mathbf{h}_k^\mathsf{H} \in \mathbb{C}^M$ denotes the kth *row* of $\mathbf{H}$. The SINR γ_k of user k reads

$$\gamma_k = \frac{|\mathbf{h}_k^\mathsf{H}\mathbf{g}_k|^2}{\sum_{j=1, j\neq k}^{M} |\mathbf{h}_k^\mathsf{H}\mathbf{g}_j|^2 + \frac{\sigma^2}{M}}. \tag{13}$$

The system sum rate R_{sum} is defined as

$$R_{\mathrm{sum}} = \sum_{k=1}^{K} \log_2\left(1 + \gamma_k\right) \quad \text{[bits/s/Hz]}. \tag{14}$$

B. Channel Model

Under the assumption of a rich scattering environment, the correlated channel can be modeled as [27]–[29]

$$\mathbf{H} = \mathbf{L}^{1/2}\mathbf{X}\mathbf{\Theta}^{1/2}, \tag{15}$$

where $\mathbf{X} \in \mathbb{C}^{K\times M}$ has i.i.d. zero-mean entries of variance $1/M$, $\mathbf{\Theta} \in \mathbb{C}^{M\times M}$ is the nonnegative definite correlation matrix at the transmitter with eigenvalues $\lambda_1, \ldots, \lambda_M$, ordered as $\lambda_1 \leq \ldots \leq \lambda_M$, and $\mathbf{L} = \mathrm{diag}(l_1, \ldots, l_K)$, with entries ordered as $l_1 \leq \ldots \leq l_K$, contains the user channel gains, i.e. the inverse user path losses. Note that in (15) the entries of $\mathbf{X}$ are not required to be Gaussian. We further assume that there exist $a_l, a_u > 0$ such that,

$$a_l < \lambda_1 \leq \lambda_M < a_u, \tag{16}$$

$$a_l < l_1 \leq l_K < a_u \tag{17}$$

uniformly on M and K. That is, (16) assumes that the correlation between transmit antennas does not increase as the number of antennas increases. For practical finite dimensional systems, this is equivalent to requesting that neighboring antennas are spaced sufficiently apart. Equation (17) assumes that the users are not too close to the base station but not too far away either; this is a realistic assumption, as distant users would be served by neighboring base stations. Those requirements, although rather realistic, are obviously not mandatory

for practical systems; however, they are required for the mathematical derivations of the present article.

Note that field measurements [30] suggest that a user-invariant correlation matrix $\boldsymbol{\Theta}$ is not a fully realistic assumption. Signal correlation at the transmitter does not only arise from close antenna spacing but also from the channel diversity and more specifically from the distribution of the solid angles of departure of effectively received energy. It could be argued though, that the scenario where all users experience equal transmit covariance matrices represents a worst case scenario, as it reduces multi-user diversity. If not fully realistic, the current assumption on $\boldsymbol{\Theta}$ is therefore still an interesting hypothesis. Further note that (15) assumes that the receivers are spaced sufficiently apart and are therefore spatially uncorrelated, an assumption which could also be argued against in some specific scenarios.

Besides, we suppose that only $\hat{\mathbf{H}}$, an imperfect estimate of the true channel matrix $\mathbf{H}$, is available at the transmitter. The channel gain matrix $\mathbf{L}$ as well as the transmit correlation $\boldsymbol{\Theta}$ are assumed to be slowly varying compared to the channel coherence time and are assumed to be perfectly known to the transmitter. We model $\hat{\mathbf{H}}$ as

$$\hat{\mathbf{H}} = \mathbf{L}^{1/2}\hat{\mathbf{X}}\boldsymbol{\Theta}^{1/2} \tag{18}$$

with imperfect short-term statistics $\hat{\mathbf{X}}$ of the form

$$\hat{\mathbf{X}} = \operatorname{diag}\left(\sqrt{1-\tau_1^2}, \dots, \sqrt{1-\tau_K^2}\right)\mathbf{X} + \operatorname{diag}\left(\tau_1, \dots, \tau_K\right)\mathbf{Q}, \tag{19}$$

where $\mathbf{Q} \in \mathbb{C}^{K\times M}$ is the matrix of channel estimation errors containing i.i.d. entries of zero mean and variance $1/M$, and $\tau_k \in [0,1]$. The parameters τ_k reflect the amount of distortion in the channel estimate $\mathbf{h}_k$ of user k. We assume that the τ_k are perfectly known at the transmitter. However, as shown in [31], an approximate knowledge of τ_k will not lead to a severe performance degradation of the system. Furthermore, we suppose that $\mathbf{X}$ and $\mathbf{Q}$ are mutually independent as well as independent of the symbol vector $\mathbf{s}$ and noise vector $\mathbf{n}$. A similar model for the imperfect CSIT has been used in [31]–[33].

Note that the assumption of not necessarily Gaussian entries of channel matrix $\hat{\mathbf{X}}$ can be useful in characterizing the aspect of quantization of the channel estimate, which, in practice, is not necessarily Gaussian.

IV. A Deterministic Equivalent of the SINR

In the following we derive a deterministic equivalent γ_k° of the SINR γ_k of user k. That is, γ_k° is an approximation of γ_k independent of the particular realizations of $\mathbf{X}, \mathbf{Q}$, and is such that,

$$\gamma_k - \gamma_k^\circ \xrightarrow{M\to\infty} 0, \tag{20}$$

almost surely. We proceed by deriving $\gamma_{k,\mathrm{rzf}}^\circ$ for RZF precoding with regularization parameter α and subsequently let $\alpha \to 0$ to obtain $\gamma_{k,\mathrm{zf}}^\circ$ for ZF precoders.

Several known results of random matrix theory (RMT) that we apply during the derivations are recalled in Appendix IV.

A. Regularized Zero-forcing Precoding

Consider the RZF precoding matrix

$$\mathbf{G}_{\mathrm{rzf}} = \sqrt{M}\xi\left(M\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + M\alpha\mathbf{I}_M\right)^{-1}\hat{\mathbf{H}}^{\mathsf{H}}, \tag{21}$$

where we remind that $\hat{\mathbf{H}}$ is the estimated channel matrix available at the transmitter and the scaling factor ξ is set to fulfill the power constraint (11). The regularization scalar $\alpha > 0$ in (21) is scaled by M to ensure that, as (K, M) grow large, both $\mathrm{tr} M\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ and $\mathrm{tr} M\alpha\mathbf{I}_M$ grow with the same order of magnitude.

Theorem 2: Let $\gamma_{k,\mathrm{rzf}}$ be the SINR of user k for RZF precoding. Then

$$\gamma_{k,\mathrm{rzf}} - \gamma_{k,\mathrm{rzf}}^\circ \xrightarrow{M\to\infty} 0, \tag{22}$$

almost surely, where $\gamma_{k,\mathrm{rzf}}^\circ$ is given by

$$\gamma_{k,\mathrm{rzf}}^\circ = \frac{l_k^2(1-\tau_k^2)\left(m^\circ\right)^2}{l_k\Upsilon^\circ(1-\tau_k^2[1-(1+l_k m^\circ)^2]) + \frac{\Psi^\circ(\alpha)}{\rho}(1+l_k m^\circ)^2} \tag{23}$$

with

$$m^\circ = \mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-1}, \tag{24}$$

$$\Psi^\circ(\alpha) = c\mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-2} - \frac{\alpha\frac{1}{\beta}\mathrm{Tr}\mathbf{L}^2\left(\mathbf{I}_K + m^\circ\mathbf{L}\right)^{-2}\left[\mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-2}\right]^2}{1 - \frac{1}{\beta}\mathrm{Tr}\mathbf{L}^2\left(\mathbf{I}_K + m^\circ\mathbf{L}\right)^{-2}\mathrm{Tr}\boldsymbol{\Theta}^2\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-2}}, \tag{25}$$

$$\Upsilon^\circ = m^\circ - \frac{\alpha\mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-2}}{1 - \frac{1}{\beta}\mathrm{Tr}\mathbf{L}^2\left(\mathbf{I}_K + m^\circ\mathbf{L}\right)^{-2}\mathrm{Tr}\boldsymbol{\Theta}^2\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-2}}, \tag{26}$$

where c is the unique real positive solution of

$$c = \frac{1}{\beta}\mathrm{Tr}\mathbf{L}\left(\mathbf{I}_K + \mathbf{L}\mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_M + c\boldsymbol{\Theta}\right)^{-1}\right)^{-1}. \tag{27}$$

Moreover, define $m_0 = 1/\alpha$, and for $k \geq 1$

$$c_k = \frac{1}{\beta}\mathrm{Tr}\mathbf{L}\left(\mathbf{I}_K + m_{k-1}\mathbf{L}\right)^{-1}, \tag{28}$$

$$m_k = \mathrm{Tr}\boldsymbol{\Theta}\left(\alpha\mathbf{I}_N + c_k\boldsymbol{\Theta}\right)^{-1}. \tag{29}$$

Then $c = \lim_{k\to\infty} c_k$.

Proof: From the sum power constraint (11) we obtain

$$\xi^2 = \frac{P}{\frac{1}{M}\mathrm{tr}\left[\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}\left(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + \alpha\mathbf{I}_M\right)^{-2}\right]} \tag{30}$$

$$\stackrel{(a)}{=} \frac{P}{m_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha) - \alpha m'_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha)} = \frac{P}{\Psi(\alpha)}, \tag{31}$$

where (a) follows from the decomposition $\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + \alpha\mathbf{I}_M)^{-2} = (\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + \alpha\mathbf{I}_M)^{-1} - \alpha(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + \alpha\mathbf{I}_M)^{-2}$, and we define

$$\Psi(\alpha) \triangleq m_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha) - \alpha m'_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha) \tag{32}$$

with $m'_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha)$ the derivative of $m_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z)$ w.r.t. z in $z = -\alpha$.

The received symbol y_k of user k is given by

$$y_k = \xi \mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{h}}_k s_k + \xi \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{h}}_i s_i + n_k, \tag{33}$$

where $\hat{\mathbf{W}} \triangleq (\hat{\mathbf{H}}^{\mathsf{H}} \hat{\mathbf{H}} + \alpha \mathbf{I}_M)^{-1}$ and $\hat{\mathbf{h}}_k^{\mathsf{H}}$ denotes the kth *row* of $\hat{\mathbf{H}}$. The SINR $\gamma_{k,\mathrm{rzf}}$ of user k can be written in the form

$$\gamma_{k,\mathrm{rzf}} = \frac{|\mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{h}}_k|^2}{\mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{H}}_{[k]}^{\mathsf{H}} \hat{\mathbf{H}}_{[k]} \hat{\mathbf{W}} \mathbf{h}_k + \frac{1}{\rho} \Psi(\alpha)}, \tag{34}$$

where $\hat{\mathbf{H}}_{[k]}^{\mathsf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_{k-1}, \hat{\mathbf{h}}_{k+1}, \ldots, \hat{\mathbf{h}}_K] \in \mathbb{C}^{M \times (K-1)}$.

We will proceed by successively deriving deterministic equivalent expressions for $\Psi(\alpha)$, for the signal power $|\mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{h}}_k|^2$ and for the power of the interference $\mathbf{h}_k^{\mathsf{H}} \hat{\mathbf{W}} \hat{\mathbf{H}}_{[k]}^{\mathsf{H}} \hat{\mathbf{H}}_{[k]} \hat{\mathbf{W}} \mathbf{h}_k$.

Consider $\Psi(\alpha)$ in (32). From Theorem 1, $m_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha)$ is close to $m^\circ_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha)$ given by (4) as

$$m^\circ_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha) = \mathrm{Tr}\left(\alpha \mathbf{I}_M + c\mathbf{\Theta}\right)^{-1}, \tag{35}$$

where c and m° are defined in (27) and (24), respectively.

Remark 2: Note that m° in (24) is a deterministic equivalent of the Stieltjes transform $m_{\mathbf{A}}(z)$ of

$$\mathbf{A} \triangleq \hat{\mathbf{X}}^{\mathsf{H}} \mathbf{L} \hat{\mathbf{X}} + \alpha \mathbf{\Theta}^{-1}, \tag{36}$$

evaluated at $z=0$, i.e. almost surely

$$m_{\mathbf{A}}(0) - m^\circ \overset{M\to\infty}{\longrightarrow} 0. \tag{37}$$

It is uncommon to consider Stieltjes transforms evaluated at $z = 0$. However, in our case this is valid, because we assumed in (16) that $1/\lambda_M > 1/a_u$ and then the smallest eigenvalue of $\mathbf{A}$ is strictly greater than $1/(2a_u) > 0$, uniformly on M. Therefore, $m_{\mathbf{A}}(0)$ is well defined. Now $m_{\mathbf{A}}(0) = m_{\mathbf{A}-1/(2a_u)\mathbf{I}_M}(-1/(2a_u))$. Since $\mathbf{A} - 1/(2a_u)\mathbf{I}_M$ meets the conditions of Theorem 1 with $\mathbf{S}_N = \alpha\mathbf{\Theta}^{-1} - 1/(2a_u)\mathbf{I}_M$, $\mathbf{T}_N = \mathbf{L}$ and $\mathbf{R}_N = \mathbf{Q}_N = \mathbf{I}_N$, one can determine a deterministic equivalent for $m_{\mathbf{A}}(0)$, which is then m°.

Since the deterministic equivalent for the Stieltjes transform of $\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ is itself the Stieltjes transform of a probability distribution (cf. [34]), a trivial application of the dominated convergence theorem ensures that the derivative $m^{\circ\prime}_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z)$ of $m^\circ_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z)$ is a deterministic equivalent for $m'_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z)$, i.e.

$$m'_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z) - m^{\circ\prime}_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(z) \overset{M\to\infty}{\longrightarrow} 0, \tag{38}$$

almost surely.

After differentiation of (35) and standard algebraic manipulations, we obtain

$$\Psi(\alpha) - \Psi^\circ(\alpha) \overset{M\to\infty}{\longrightarrow} 0, \tag{39}$$

almost surely, where

$$\Psi^\circ(\alpha) \triangleq m^\circ_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha) - \alpha m^{\circ\prime}_{\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}}(-\alpha), \tag{40}$$

which is explicitly given by (25).

1) A Deterministic Equivalent Of The Signal Power: Applying Lemma 4 to $\hat{\mathbf{h}}_k^{\mathsf{H}} \hat{\mathbf{W}} = \hat{\mathbf{h}}_k^{\mathsf{H}} (\hat{\mathbf{H}}_{[k]}^{\mathsf{H}} \hat{\mathbf{H}}_{[k]} + \alpha \mathbf{I}_M + \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathsf{H}})^{-1}$, we have

$$\hat{\mathbf{h}}_k^{\mathsf{H}} \hat{\mathbf{W}} \mathbf{h}_k = \frac{\hat{\mathbf{h}}_k^{\mathsf{H}} \left(\hat{\mathbf{H}}_{[k]}^{\mathsf{H}} \hat{\mathbf{H}}_{[k]} + \alpha \mathbf{I}_M\right)^{-1} \mathbf{h}_k}{1 + \hat{\mathbf{h}}_k^{\mathsf{H}} \left(\hat{\mathbf{H}}_{[k]}^{\mathsf{H}} \hat{\mathbf{H}}_{[k]} + \alpha \mathbf{I}_M\right)^{-1} \hat{\mathbf{h}}_k}. \tag{41}$$

Together with $\hat{\mathbf{h}}_k^{\mathsf{H}} = \sqrt{l_k}(\sqrt{1-\tau_k^2}\mathbf{x}_k^{\mathsf{H}} + \tau_k \mathbf{q}_k^{\mathsf{H}})\mathbf{\Theta}^{1/2}$ we obtain

$$\hat{\mathbf{h}}_k^{\mathsf{H}} \hat{\mathbf{W}} \mathbf{h}_k = \frac{\sqrt{1-\tau_k^2}\, l_k \mathbf{x}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \mathbf{x}_k}{1 + l_k \hat{\mathbf{x}}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \hat{\mathbf{x}}_k} + \frac{\tau_k l_k \mathbf{q}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \mathbf{x}_k}{1 + l_k \hat{\mathbf{x}}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \hat{\mathbf{x}}_k}$$

with $\mathbf{A}_{[k]} = \hat{\mathbf{X}}_{[k]}^{\mathsf{H}} \mathbf{L}_{[k]} \hat{\mathbf{X}}_{[k]} + \alpha \mathbf{\Theta}^{-1}$ for $\hat{\mathbf{X}}_{[k]}^{\mathsf{H}} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{k-1}, \hat{\mathbf{x}}_{k+1}, \ldots, \hat{\mathbf{x}}_K]$, $\hat{\mathbf{x}}_n$ being the nth *row* of $\hat{\mathbf{X}}$, and $\mathbf{L}_{[k]} = \mathrm{diag}(l_1, \ldots, l_{k-1}, l_{k+1} \ldots l_K)$. Since both $\hat{\mathbf{x}}_k$ and $\mathbf{x}_k$ have i.i.d. entries of variance $1/M$ and are independent of $\mathbf{A}_{[k]}$ we evoke Corollary 7 and obtain, almost surely,

$$\mathbf{x}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \mathbf{x}_k - \frac{1}{M} \mathrm{tr} \mathbf{A}_{[k]}^{-1} \overset{M\to\infty}{\longrightarrow} 0, \tag{42}$$

$$\hat{\mathbf{x}}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \hat{\mathbf{x}}_k - \frac{1}{M} \mathrm{tr} \mathbf{A}_{[k]}^{-1} \overset{M\to\infty}{\longrightarrow} 0. \tag{43}$$

Similarly, as $\mathbf{q}_k$ and $\mathbf{x}_k$ are independent, it follows from Lemma 6 that

$$\mathbf{q}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \mathbf{x}_k \overset{M\to\infty}{\longrightarrow} 0, \tag{44}$$

almost surely. Consequently, since $(1 + l_k \hat{\mathbf{x}}_k^{\mathsf{H}} \mathbf{A}_{[k]}^{-1} \hat{\mathbf{x}}_k)$ is bounded away from zero, we obtain

$$\hat{\mathbf{h}}_k^{\mathsf{H}} \hat{\mathbf{W}} \mathbf{h}_k - \sqrt{1-\tau_k^2} \frac{l_k \frac{1}{M} \mathrm{tr} \mathbf{A}_{[k]}^{-1}}{1 + l_k \frac{1}{M} \mathrm{tr} \mathbf{A}_{[k]}^{-1}} \overset{M\to\infty}{\longrightarrow} 0, \tag{45}$$

almost surely.

We require Lemma 7 to prove that a rank-1 perturbation of $\mathbf{A}$ has no impact on $\mathrm{Tr}\mathbf{A}^{-1}$ for asymptotically large M. Rewrite $\mathbf{A}_{[k]}^{-1}$ as,

$$\mathbf{A}_{[k]}^{-1} = \left(\hat{\mathbf{X}}_{[k]}^{\mathsf{H}} \mathbf{L}_{[k]} \hat{\mathbf{X}}_{[k]} + \alpha \mathbf{\Theta}^{-1}\right)^{-1} = \left(\left[\hat{\mathbf{X}}_{[k]}^{\mathsf{H}} \mathbf{L}_{[k]} \hat{\mathbf{X}}_{[k]} + \alpha \mathbf{\Theta}^{-1} - \alpha \frac{1}{2a_u} \mathbf{I}_M\right] + \alpha \frac{1}{2a_u} \mathbf{I}_M\right)^{-1}. \tag{46}$$

Since $1/\lambda_M > 1/a_u$ uniformly on M, notice that the matrix in brackets on the right-hand side is still nonnegative definite. Thus, applying Lemma 7 to this matrix and the scalar $\alpha \frac{1}{2a_u} > 0$, we obtain

$$\frac{1}{M} \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{[k]}^{\mathsf{H}} \mathbf{L}_{[k]} \hat{\mathbf{X}}_{[k]} + \alpha \mathbf{\Theta}^{-1}\right)^{-1}\right] - \frac{1}{M} \mathrm{tr}\left[\left(\hat{\mathbf{X}}_{[k]}^{\mathsf{H}} \mathbf{L}_{[k]} \hat{\mathbf{X}}_{[k]} + l_k \hat{\mathbf{x}}_k \hat{\mathbf{x}}_k^{\mathsf{H}} + \alpha \mathbf{\Theta}^{-1}\right)^{-1}\right] \overset{M\to\infty}{\longrightarrow} 0.$$

Therefore, surely,

$$\frac{1}{M} \mathrm{tr} \mathbf{A}_{[k]}^{-1} - \frac{1}{M} \mathrm{tr} \mathbf{A}^{-1} \overset{M\to\infty}{\longrightarrow} 0, \tag{47}$$

where we remind that $\mathbf{A} = \hat{\mathbf{X}}^{\mathsf{H}} \mathbf{L} \hat{\mathbf{X}} + \alpha \mathbf{\Theta}^{-1}$.

Thus, (47) and (37) imply

$$\frac{1}{M}\mathrm{tr}\mathbf{A}_{[k]}^{-1} - m^\circ \overset{M\to\infty}{\longrightarrow} 0, \tag{48}$$

almost surely. Finally, (45) takes the form

$$\hat{\mathbf{h}}_k^{\mathsf{H}}\hat{\mathbf{W}}\mathbf{h}_k - \sqrt{1-\tau_k^2}\frac{l_k m^\circ}{1+l_k m^\circ} \overset{M\to\infty}{\longrightarrow} 0. \tag{49}$$

2) A Deterministic Equivalent Of The Interference Power: With $\hat{\mathbf{W}} = \mathbf{\Theta}^{-1/2}\mathbf{A}^{-1}\mathbf{\Theta}^{-1/2}$, the interference power can be written as

$$\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k = l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\hat{\mathbf{X}}_{[k]}^{\mathsf{H}}\mathbf{L}_{[k]}\hat{\mathbf{X}}_{[k]}\mathbf{A}^{-1}\mathbf{x}_k. \tag{50}$$

Denote $c_0 = (1-\tau_k^2)l_k$, $c_1 = \tau_k^2 l_k$ and $c_2 = \tau_k\sqrt{1-\tau_k^2}l_k$, then

$$\mathbf{A} = \mathbf{A}_{[k]} + c_0\mathbf{x}_k\mathbf{x}_k^{\mathsf{H}} + c_1\mathbf{q}_k\mathbf{q}_k^{\mathsf{H}} + c_2\mathbf{x}_k\mathbf{q}_k^{\mathsf{H}} + c_2\mathbf{q}_k\mathbf{x}_k^{\mathsf{H}}. \tag{51}$$

In order to eliminate the dependence between $\mathbf{x}_k$ and $\mathbf{A}$ in (50), we rewrite (50) as

$$\begin{aligned}\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k &= l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\hat{\mathbf{X}}_{[k]}^{\mathsf{H}}\mathbf{L}_{[k]}\hat{\mathbf{X}}_{[k]}\mathbf{A}^{-1}\mathbf{x}_k \\ &+ l_k\mathbf{x}_k^{\mathsf{H}}\left[\mathbf{A}^{-1} - \mathbf{A}_{[k]}^{-1}\right]\hat{\mathbf{X}}_{[k]}^{\mathsf{H}}\mathbf{L}_{[k]}\hat{\mathbf{X}}_{[k]}\mathbf{A}^{-1}\mathbf{x}_k. \end{aligned}\tag{52}$$

Applying Lemma 8 to the term in brackets in (52) and together with $\mathbf{A}_{[k],k}^{-1}\hat{\mathbf{X}}_{[k]}^{\mathsf{H}}\mathbf{L}_{[k]}\hat{\mathbf{X}}_{[k]} = \mathbf{I}_M - \alpha\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}$, equation (52) takes the form

$$\begin{aligned}\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k &= l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k - \alpha l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}\mathbf{A}^{-1}\mathbf{x}_k \\ &- c_0 l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k\left(\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k - \alpha\mathbf{x}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}\mathbf{A}^{-1}\mathbf{x}_k\right) \\ &- c_1 l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{q}_k\left(\mathbf{q}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k - \alpha\mathbf{q}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}\mathbf{A}^{-1}\mathbf{x}_k\right) \\ &- c_2 l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k\left(\mathbf{q}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k - \alpha\mathbf{q}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}\mathbf{A}^{-1}\mathbf{x}_k\right) \\ &- c_2 l_k\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{q}_k\left(\mathbf{x}_k^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}_k - \alpha\mathbf{x}_k^{\mathsf{H}}\mathbf{A}_{[k]}^{-1}\mathbf{\Theta}^{-1}\mathbf{A}^{-1}\mathbf{x}_k\right). \end{aligned}\tag{53}$$

To find a deterministic equivalent for all of the 14 quadratic terms in (53) we need the following lemma, which is an extension of Corollary 7.

Lemma 1: Let $\mathbf{U}, \mathbf{V} \in \mathbb{C}^{N\times N}$ be invertible and of uniformly bounded spectral norm. Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and finite 8th order moment and be mutually independent as well as independent of $\mathbf{U}, \mathbf{V}$. Define $c_0, c_1, c_2 \in \mathbb{R}^+$ such that $c_0c_1 - c_2^2 \geq 0$ and let $u \triangleq \frac{1}{N}\mathrm{tr}\mathbf{V}^{-1}$ and $u' \triangleq \frac{1}{N}\mathrm{tr}\mathbf{U}\mathbf{V}^{-1}$. Then we have

$$\begin{aligned}&\mathbf{x}^{\mathsf{H}}\mathbf{U}\left(\mathbf{V} + c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}}\right)^{-1}\mathbf{x} \\ &- \frac{u'(1+c_1u)}{(c_0c_1 - c_2^2)u^2 + (c_0+c_1)u + 1} \overset{N\to\infty}{\longrightarrow} 0,\end{aligned} \tag{54}$$

almost surely. Furthermore

$$\begin{aligned}&\mathbf{x}^{\mathsf{H}}\mathbf{U}\left(\mathbf{V} + c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}}\right)^{-1}\mathbf{y} \\ &- \frac{-c_2uu'}{(c_0c_1 - c_2^2)u^2 + (c_0+c_1)u + 1} \overset{N\to\infty}{\longrightarrow} 0,\end{aligned} \tag{55}$$

almost surely.

The proof of Lemma 1 is left to Appendix III.

Denote $u \triangleq \mathrm{Tr}\mathbf{A}_{[k]}^{-1}$ and $u' \triangleq \mathrm{Tr}\mathbf{\Theta}^{-1}\mathbf{A}_{[k]}^{-2}$. Notice that in our case $c_0c_1 = c_2^2$, thus $(c_0c_1 - c_2^2)u^2 + (c_0+c_1)u + 1$ reduces to $1 + l_ku$. Applying Lemma 1 to each of the 14 terms in (53) we obtain

$$\begin{aligned}&\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k - \\ &\left[\frac{l_k(1+c_1u)(u-\alpha u')}{1+l_ku} - \frac{l_kc_0u(u-\alpha u')}{(1+l_ku)^2}\right] \overset{M\to\infty}{\longrightarrow} 0\end{aligned} \tag{56}$$

almost surely, where the first term in brackets stems from the first line in (53) and the second term in brackets of (56) arises from the last four lines of equation (53). Replacing c_0 and c_1 by $(1-\tau_k^2)l_k$ and $\tau_k^2 l_k$, respectively and after some algebraic manipulation, (56) takes the form

$$\begin{aligned}&\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k - \\ &\frac{l_k(u-\alpha u')\left[1-\tau_k^2\left(1-(1+l_ku)^2\right)\right]}{(1+l_ku)^2} \overset{M\to\infty}{\longrightarrow} 0,\end{aligned} \tag{57}$$

almost surely. Since a rank-1 perturbation has no impact on $\mathrm{Tr}\mathbf{A}^{-1}$ for $M\to\infty$, we surely have

$$u - m_{\mathbf{A}}(0) \overset{M\to\infty}{\longrightarrow} 0, \tag{58}$$

$$u' - m_{\mathbf{A}^2,\mathbf{\Theta}^{-1}}(0) \overset{M\to\infty}{\longrightarrow} 0. \tag{59}$$

The definition of $m_{\mathbf{A}^2,\mathbf{\Theta}^{-1}}(z)$ can be extended to $z=0$, since $\mathbf{A}$ and $\mathbf{\Theta}$ have their largest eigenvalue belonging to a compact set away from zero, uniformly on M (cf. Remark 2). Denote

$$\Upsilon = m_{\mathbf{A}}(0) - \alpha m'_{\mathbf{A},\mathbf{\Theta}^{-1}}(0). \tag{60}$$

Observe that

$$m_{\mathbf{A}^2,\mathbf{\Theta}^{-1}}(0) = m'_{\mathbf{A},\mathbf{\Theta}^{-1}}(0). \tag{61}$$

Furthermore, we have $m_{\mathbf{A}^2,\mathbf{\Theta}^{-1}}(0) - m^{\circ\prime}_{\mathbf{A},\mathbf{\Theta}^{-1}}(0) \overset{M\to\infty}{\longrightarrow} 0$, almost surely, where $m^{\circ\prime}_{\mathbf{A},\mathbf{\Theta}^{-1}}(0)$ is given by Theorem 1. Similar to the derivations of $\Psi^\circ(\alpha)$ leading to (25), we then obtain that $\Upsilon - \Upsilon^\circ \overset{M\to\infty}{\longrightarrow} 0$, almost surely, with Υ° given by

$$\Upsilon^\circ = m^\circ_{\mathbf{A}}(0) - \alpha m^{\circ\prime}_{\mathbf{A},\mathbf{\Theta}^{-1}}(0), \tag{62}$$

whose explicit form is given by (26). Substituting u and u' in (57) by their respective deterministic equivalent expressions m° and $m^{\circ\prime}_{\mathbf{A}\mathbf{\Theta}^{-1}}(0)$, we obtain

$$\begin{aligned}&\mathbf{h}_k^{\mathsf{H}}\hat{\mathbf{W}}\hat{\mathbf{H}}_{[k]}^{\mathsf{H}}\hat{\mathbf{H}}_{[k]}\hat{\mathbf{W}}\mathbf{h}_k - \\ &\frac{l_k\Upsilon^\circ(1-\tau_k^2[1-(1+l_km^\circ)^2])}{(1+l_km^\circ)^2} \overset{M\to\infty}{\longrightarrow} 0,\end{aligned} \tag{63}$$

almost surely.

Finally, a deterministic equivalent $\gamma^\circ_{k,\mathrm{rzf}}$ of $\gamma_{k,\mathrm{rzf}}$ is given by (23), since the denominator of (23) is positive and bounded away from zero.

Moreover, the convergence of the sequence $\{c_k\}$ to c is a direct consequence of Proposition 1. ■

B. Optimal Regularized Zero-forcing Precoding

Define $R^{\circ,\mathrm{rzf}}_{\mathrm{sum}}$ to be

$$R^{\circ,\mathrm{rzf}}_{\mathrm{sum}} = \sum_{k=1}^{K}\log_2\left(1+\gamma^\circ_{k,\mathrm{rzf}}\right) \quad [\text{bits/s/Hz}], \tag{64}$$

where the parameter α in $\gamma^{\circ}_{k,\mathrm{rzf}}$ is chosen to be the positive real that maximizes $R^{\circ,\mathrm{rzf}}_{\mathrm{sum}}$, i.e. $\alpha = \alpha^{\star\circ}$, with $\alpha^{\star\circ}$ defined as

$$\alpha^{\star\circ} = \arg\max_{\alpha>0} \left\{ R^{\circ,\mathrm{rzf}}_{\mathrm{sum}} \right\}. \tag{65}$$

For the general channel model (18), $\alpha^{\star\circ}$ is a solution, s.t. $\alpha^{\star\circ} > 0$, of the implicit equation

$$\sum_{k=1}^{K} \frac{\partial \gamma^{\circ}_{k,\mathrm{rzf}}}{\partial \alpha} \frac{1}{1+\gamma^{\circ}_{k,\mathrm{rzf}}} = 0. \tag{66}$$

The implicit equation (66) is not convex in α and the solution can be computed via a one-dimensional line search[1]. Note that for non i.i.d. channels the RZF in (21) is not asymptotically optimal anymore.

The RZF precoder with optimal regularization parameter $\alpha^{\star\circ}$ is called ORZF. For homogeneous networks ($\mathbf{L} = \mathbf{I}_K$) the user channels $\mathbf{h}_k$ are statistically equivalent and it is reasonable to assume that the distortions τ_k^2 of the CSIT $\hat{\mathbf{h}}_k$ are identical for all users, i.e. $\tau \triangleq \tau_k$. Under the additional assumption of uncorrelated transmit antennas ($\mathbf{\Theta} = \mathbf{I}_M$) , the solution to (65) has a closed form and leads to the asymptotically optimal precoder [17].

Proposition 2: Let $\mathbf{\Theta} = \mathbf{I}_M$ and $\mathbf{L} = \mathbf{I}_K$. The approximated SINR $\gamma^{\circ}_{k,\mathrm{rzf}}$ of user k under RZF precoding (equivalently, the approximated per-user rate and the sum rate) is maximized for a regularization term $\alpha \triangleq \alpha^{\star\circ}$, given by

$$\alpha^{\star\circ} = \left(\frac{1+\tau^2\rho}{1-\tau^2} \right) \frac{1}{\beta\rho}. \tag{67}$$

Proof: For $\mathbf{L} = \mathbf{I}_K$ and $\mathbf{\Theta} = \mathbf{I}_M$, $\Psi^{\circ}(\alpha) = \Upsilon^{\circ}$ and $m^{\circ} = m^{\circ}_{\hat{\mathbf{X}}^{\mathsf{H}}\hat{\mathbf{X}}}(-\alpha)$ is the Stieltjes transform of the Marčenko-Pastur law and reads [24]

$$m^{\circ}_{\hat{\mathbf{X}}^{\mathsf{H}}\hat{\mathbf{X}}}(-\alpha) = \frac{\beta(1-\alpha) - 1 + d(\alpha,\beta)}{2\alpha\beta}$$
$$\text{with} \quad d(\alpha,\beta) = \sqrt{\beta^2\alpha^2 + 2\alpha\beta(1+\beta) + (1-\beta)^2}. \tag{68}$$

Substituting (68) into (23)-(27) and setting the derivative w.r.t. α to zero, a real positive solution is given by (67). ■

Notice that for perfect CSIT ($\tau = 0$) we have $\alpha^{\star\circ} = 1/(\beta\rho)$ which corresponds to the RZF-CDU precoder derived in [9], [18]. As mentioned in [9], for *large* (K, M) the RZF-CDU precoder is identical to the MMSE precoder in [16], [31]. In contrast, for $\tau > 0$, the ORZF transmit filter and the MMSE transmit filter [31] are not identical anymore, even in the large M limit. Furthermore, for $\tau > 0$ at asymptotically high SNR the regularization term $\alpha^{\star\circ}$ in (67) converges to

$$\lim_{\rho\to\infty} \alpha^{\star\circ} = \frac{\tau^2}{1-\tau^2} \frac{1}{\beta}. \tag{69}$$

Thus, for asymptotically high SNR, ORZF does *not* converge to ZF precoding, in the sense that $\alpha^{\star\circ}$ does not converge to 0. Also note that for $\beta \to \infty$, ORZF converges to ZF regardless of τ^2. The same has been observed in [31] for the MMSE precoder.

With (67), the SINR (23) takes the following simplified form.

[1]However, in simulations we observe only a single maximum for $\alpha > 0$. In this case $\alpha^{\star\circ}$ can be computed very efficiently.

Corollary 1: Let $\mathbf{\Theta} = \mathbf{I}_M$, $\mathbf{L} = \mathbf{I}_K$ and $\tau_k = \tau$ and $\gamma_{k,\mathrm{rzf}}$ be the SINR of user k under ORZF precoding. Then

$$\gamma_{k,\mathrm{rzf}} - \gamma^{\circ}_{k,\mathrm{rzf}} \xrightarrow{M\to\infty} 0, \tag{70}$$

almost surely, where $\gamma^{\circ}_{k,\mathrm{rzf}}$ is given by

$$\gamma^{\circ}_{k,\mathrm{rzf}} \triangleq \gamma^{\circ}_{\mathrm{rzf}} = m^{\circ}_{\hat{\mathbf{X}}^{\mathsf{H}}\hat{\mathbf{X}}}(-\alpha^{\star\circ}) = \frac{\omega}{2}\rho(\beta-1) + \frac{\chi}{2} - \frac{1}{2}, \tag{71}$$

where $\omega \in [0,1]$ and χ are given by

$$\omega = \frac{1-\tau^2}{1+\tau^2\rho}, \tag{72}$$
$$\chi = \sqrt{(\beta-1)^2\omega^2\rho^2 + 2(1+\beta)\omega\rho + 1}. \tag{73}$$

Proof: Replace α in (23) by $\alpha^{\star\circ}$ in (67). After some algebraic manipulations we obtain (71). ■

Note that for $\tau^2 = 0$ and $\alpha = \alpha^{\star\circ}$, (71) is identical to the asymptotic SINR derived in [18] and for the inter-cell interference-free system in [19].

For $\beta = 1$, equation (71) simplifies to

$$\gamma^{\circ}_{k,\mathrm{rzf}} = -\frac{1}{2} + \sqrt{\omega\rho + \frac{1}{4}}. \tag{74}$$

Note that, as shown in the following subsection, $\omega\rho(\beta-1)$ turns out to be a deterministic equivalent for the SINR of ZF precoding for $\beta > 1$. Let $\Delta\gamma^{\circ}$ be the SINR gap between ORZF and ZF precoding

$$\Delta\gamma^{\circ} = -\frac{\omega}{2}\rho(\beta-1) + \frac{\chi}{2} - \frac{1}{2}. \tag{75}$$

Then for asymptotically high SNR, $\Delta\gamma^{\circ}$ converges to the following limits

$$\lim_{\rho\to\infty} \Delta\gamma^{\circ} = \begin{cases} \frac{1}{\beta-1} & \text{if } \tau^2 = 0 \\ \frac{\bar{\chi}}{2} - \frac{1-\tau^2}{\tau^2}(\beta-1) - \frac{1}{2} & \text{if } \tau^2 \neq 0, \end{cases} \tag{76}$$

where $\bar{\chi}$ is given by

$$\bar{\chi} = \sqrt{(\beta-1)^2\frac{1-\tau^2}{\tau^2} + 2(1+\beta)\frac{1-\tau^2}{\tau^2} + 1}. \tag{77}$$

Note that the SINR gap $1/(\beta-1)$ for $\tau^2 = 0$ is identical to the normalization scalar ξ^2 for ZF for large (K, M). From (76) we conclude, that ZF and ORZF do not achieve the same sum-rate at asymptotically high SNR.

C. Zero-forcing Precoding

For $\alpha = 0$, the RZF precoding matrix in (21) reduces to the ZF precoding matrix $\mathbf{G}_{\mathrm{zf}}$ which reads

$$\mathbf{G}_{\mathrm{zf}} = \frac{\xi}{\sqrt{M}} \hat{\mathbf{H}}^{\mathsf{H}} \left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}} \right)^{-1}, \tag{78}$$

where ξ is a scaling factor to fulfill the power constraint (11).

To derive a deterministic equivalent of the SINR of ZF precoding, we cannot apply the same techniques as for RZF, since by removing a row of $\hat{\mathbf{H}}$, the matrix $\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}}$ becomes singular. Therefore, we adopt a different strategy and derive the SINR $\gamma_{k,\mathrm{zf}}$ for ZF of user k and $\beta > 1$ as

$$\gamma_{k,\mathrm{zf}} = \lim_{\alpha\to 0} \gamma_{k,\mathrm{rzf}}. \tag{79}$$

The result is summarized in the following theorem.

Theorem 3: Let $\beta > 1$ and $\gamma_{k,\mathrm{zf}}$ be the SINR of user k for ZF precoding. Then

$$\gamma_{k,\mathrm{zf}} - \gamma^{\circ}_{k,\mathrm{zf}} \xrightarrow{M\to\infty} 0, \tag{80}$$

almost surely, where $\gamma^{\circ}_{k,\mathrm{zf}}$ is given by

$$\gamma^{\circ}_{k,\mathrm{zf}} = \frac{1-\tau_k^2}{l_k\tau_k^2\bar{\Upsilon}^{\circ} + \frac{\bar{\Psi}^{\circ}}{\rho}} \tag{81}$$

with

$$\bar{\Psi}^{\circ} = \frac{1}{\beta\bar{c}}\mathrm{Tr}\mathbf{L}^{-1}, \tag{82}$$

$$\bar{\Upsilon}^{\circ} = \frac{c_2/\bar{c}^2}{\beta - c_2/\bar{c}^2}\mathrm{Tr}\mathbf{L}^{-1}, \tag{83}$$

$$c_2 = \mathrm{Tr}\mathbf{\Theta}^2\left(\mathbf{I}_M + \frac{1}{\bar{c}\beta}\mathbf{\Theta}\right)^{-2} \tag{84}$$

where $\bar{c}$ is the unique solution of

$$\bar{c} = \mathrm{Tr}\mathbf{\Theta}\left(\mathbf{I}_M + \frac{1}{\bar{c}\beta}\mathbf{\Theta}\right)^{-1}. \tag{85}$$

Moreover, $\bar{c} = \lim_{k\to\infty}\bar{c}_k$, where $\bar{c}_0 = 1$ and, for $k \geq 1$,

$$\bar{c}_k = \mathrm{Tr}\mathbf{\Theta}\left(\mathbf{I}_M + \frac{1}{\bar{c}_{k-1}\beta}\mathbf{\Theta}\right)^{-1}. \tag{86}$$

Note, that by Jensen's inequality $c_2/\bar{c}^2 \geq 1$ with equality if $\mathbf{\Theta} = \mathbf{I}_M$.

Corollary 2: Let $\mathbf{\Theta} = \mathbf{I}_M$, $\mathbf{L} = \mathbf{I}_K$ and $\tau_k = \tau$ then $\gamma^{\circ}_{k,\mathrm{zf}}$ takes the explicit form

$$\gamma^{\circ}_{k,\mathrm{zf}} \triangleq \gamma^{\circ}_{\mathrm{zf}} = \omega\rho(\beta-1) = \frac{1-\tau^2}{\tau^2+\frac{1}{\rho}}(\beta-1). \tag{87}$$

Proof of Corollary 2: By substituting $\mathbf{\Theta} = \mathbf{I}_M$ and $\mathbf{L} = \mathbf{I}_K$ into (85), $\bar{c}$ is explicitly given by $\bar{c} = (\beta-1)/\beta$. Since $c_2/\bar{c}^2 = 1$ we have $\bar{\Psi}^{\circ} = \bar{\Upsilon}^{\circ} = 1/(\beta-1)$. ■

Proof of Theorem 3: Recall the terms in the SINR of RZF that depend on α, i.e. $m_{\mathbf{A}}$, Ψ and Υ

$$m_{\mathbf{A}} = \frac{1}{M}\mathrm{tr}\mathbf{\Theta}\mathbf{F} \tag{88}$$

$$\Psi = \frac{1}{M}\mathrm{tr}\left(\mathbf{F} - \alpha\mathbf{F}^2\right) \tag{89}$$

$$\Upsilon = \frac{1}{M}\mathrm{tr}\mathbf{\Theta}\left(\mathbf{F} - \alpha\mathbf{F}^2\right), \tag{90}$$

where

$$\mathbf{F} = \left(\hat{\mathbf{H}}^{\mathsf{H}}\hat{\mathbf{H}} + \alpha\mathbf{I}_M\right)^{-1} \tag{91}$$

In order to take the limit $\alpha \to 0$ of Ψ and Υ, we apply the matrix inversion lemma (MIL) to $\mathbf{F}$ in (91) and obtain

$$\mathbf{F} - \alpha\mathbf{F}^2 = \hat{\mathbf{H}}^{\mathsf{H}}\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}} + \alpha\mathbf{I}_M\right)^{-2}\hat{\mathbf{H}}, \tag{92}$$

Since $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}$ is non-singular with probability one, we can take the limit $\alpha \to 0$ of Ψ and Υ, for such $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}$,

$$\bar{\Psi} \triangleq \lim_{\alpha\to 0}\Psi = \frac{1}{M}\mathrm{tr}\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\right)^{-1} \tag{93}$$

$$\bar{\Upsilon} \triangleq \lim_{\alpha\to 0}\Upsilon = \frac{1}{M}\mathrm{tr}\hat{\mathbf{H}}\mathbf{\Theta}\hat{\mathbf{H}}^{\mathsf{H}}\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}\right)^{-2}. \tag{94}$$

Note that it is necessary to assume that $\beta > 1$ to assure that the maximum eigenvalue of matrix $(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}})^{-1}$ is bounded for all large M, almost surely. Since $m_{\mathbf{A}}$ grows with α as $O(1/\alpha)$ we have

$$\gamma_{k,\mathrm{zf}} = \lim_{\alpha\to 0}\gamma_{k,\mathrm{rzf}} = \frac{1-\tau_k^2}{l_k\tau_k^2\bar{\Upsilon} + \frac{\bar{\Psi}}{\rho}}. \tag{95}$$

Now we derive deterministic equivalents $\bar{\Psi}^{\circ}$ and $\bar{\Upsilon}^{\circ}$ for $\bar{\Psi}$ and $\bar{\Upsilon}$, respectively.

With $\mathbf{S}_N = 0$, Theorem 1 can be directly applied to find $\bar{\Psi}^{\circ}$ s.t. $\bar{\Psi} - \bar{\Psi}^{\circ} \xrightarrow{M\to\infty} 0$, almost surely, as

$$\bar{\Psi}^{\circ} = \frac{1}{\beta}m^{\circ}_{\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}}(0) = \frac{1}{\beta\bar{c}}\mathrm{Tr}\mathbf{L}^{-1}, \tag{96}$$

where $\bar{c}$ is defined in (85).

In order to find $\bar{\Upsilon}^{\circ}$, notice that we can diagonalize $\mathbf{\Theta}$ in (94) s.t. $\mathbf{\Theta} = \mathbf{U}\mathrm{diag}(\lambda_1,\ldots,\lambda_M)\mathbf{U}^{\mathsf{H}}$, where $\mathbf{U}$ is a unitary matrix, and still have i.i.d. elements in the kth column $\hat{\mathbf{x}}'_k$ of $\hat{\mathbf{X}}\mathbf{U}$. Denoting $\mathbf{C} = \hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathsf{H}}$, $\mathbf{C}_{[k]} = \hat{\mathbf{H}}_{[k]}\hat{\mathbf{H}}^{\mathsf{H}}_{[k]} - \lambda_k\mathbf{L}^{1/2}\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^{\mathsf{H}}_k\mathbf{L}^{1/2}$ and applying Lemma 4 twice, equation (94) takes the form

$$\bar{\Upsilon} = \frac{1}{M}\sum_{k=1}^{M}\lambda_k^2\frac{\hat{\mathbf{x}}'^{\mathsf{H}}_k\mathbf{L}^{1/2}\mathbf{C}^{-2}_{[k]}\mathbf{L}^{1/2}\hat{\mathbf{x}}'_k}{(1+\lambda_k\hat{\mathbf{x}}'^{\mathsf{H}}_k\mathbf{L}^{1/2}\mathbf{C}^{-1}_{[k]}\mathbf{L}^{1/2}\hat{\mathbf{x}}'_k)^2}. \tag{97}$$

Notice that $\mathbf{C}^{-1}_{[k]}$ does not have uniformly bounded spectral norm. Therefore, Lemma 5 can not be applied straightforwardly. However, since β is uniformly greater than $1+\epsilon$ for some $\epsilon > 0$, $\mathbf{C}^{-1}_{[k]}$ has almost surely bounded spectral norm for all large M. This is sufficient for the following lemma to hold.

Lemma 2: Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N \in \mathbb{C}^{N\times N}$, be a series of random matrices generated by the probability space $(\Omega, \mathcal{F}, P)$ such that, for $\omega \in A \subset \Omega$, with $P(A) = 1$, $\|\mathbf{A}_N(\omega)\| < K(\omega) < \infty$, uniformly on N. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be random vectors of i.i.d. entries such that $\mathbf{x}_N \in \mathbb{C}^N$ has entries of zero mean, variance $1/N$ and finite eighth order moment, independent of $\mathbf{A}_N$. Then

$$\mathbf{x}^{\mathsf{H}}_N\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\mathrm{tr}\mathbf{A}_N \xrightarrow{N\to\infty} 0,$$

almost surely.

The proof of Lemma 2 is left to Appendix V.

Moreover, the assumptions of the rank-1 perturbation Lemma 7 are no longer satisfied. Thus, it cannot be ensured that the normalized trace of $\mathbf{C}^{-1}_{[k]}$ and $\mathbf{C}^{-1}$ are asymptotically equal. In fact, following the same line of argument as above, we also have a generalized rank-1 perturbation lemma, which now holds only almost surely.

Lemma 3: Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N \in \mathbb{C}^{N\times N}$ be deterministic with uniformly bounded spectral norm and $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_N \in \mathbb{C}^{N\times N}$ be random Hermitian, with eigenvalues $\lambda_1^{\mathbf{B}_N} \leq \ldots \leq \lambda_N^{\mathbf{B}_N}$, generated by the probability space $(\Omega, \mathcal{F}, P)$, such that, for $\omega \in B$, with $P(B) = 1$, $\lambda_1^{\mathbf{B}_N(\omega)} > \varepsilon(\omega)$, uniformly on N. Then for $\mathbf{v} \in \mathbb{C}^N$,

$$\frac{1}{N}\mathrm{tr}\mathbf{A}_N\mathbf{B}_N^{-1} - \frac{1}{N}\mathrm{tr}\mathbf{A}_N(\mathbf{B}_N + \mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1} \xrightarrow{N\to\infty} 0,$$

almost surely, where $\mathbf{B}_N^{-1}$ and $(\mathbf{B}_N + \mathbf{v}\mathbf{v}^{\mathsf{H}})^{-1}$ exist with probability one.

The proof of Lemma 3 is left to Appendix VI.

Applying Lemma 2 together with the rank-1 perturbation Lemma 3, we obtain

$$\bar{\Upsilon} - \frac{1}{\beta}\mathrm{Tr}\mathbf{L}\mathbf{C}^{-2}\frac{1}{M}\sum_{k=1}^{M}\frac{\lambda_k^2}{(1+\lambda_k\frac{1}{\beta}\mathrm{Tr}\mathbf{L}\mathbf{C}^{-1})^2} \xrightarrow{M\to\infty} 0, \quad (98)$$

almost surely. To determine a deterministic equivalent $m^\circ_{\mathbf{C},\mathbf{L}}(0)$ for $m_{\mathbf{C},\mathbf{L}}(0) = \mathrm{Tr}\mathbf{L}\mathbf{C}^{-1}$, we apply Theorem 1 as for (96) (again, the definition of $m_{\mathbf{C},\mathbf{L}}(z)$ can be easily extended to $z=0$). For $\mathrm{Tr}\mathbf{L}\mathbf{C}^{-2}$ we have

$$\mathrm{Tr}\mathbf{L}\mathbf{C}^{-2} = m_{\mathbf{C}^2,\mathbf{L}}(z) = m'_{\mathbf{C},\mathbf{L}}(0). \quad (99)$$

The derivative of $m^\circ_{\mathbf{C},\mathbf{L}}(0)$ is a deterministic equivalent of $m'_{\mathbf{C},\mathbf{L}}(0)$, so that applied to (98), we have $\bar{\Upsilon}^\circ$, s.t. $\bar{\Upsilon} - \bar{\Upsilon}^\circ \xrightarrow{M\to\infty} 0$, almost surely, that satisfies

$$\bar{\Upsilon}^\circ = \frac{c_2/\bar{c}^2}{\beta - c_2/\bar{c}^2}\mathrm{Tr}\mathbf{L}^{-1}, \quad (100)$$

where $\bar{c}$ and c_2 are defined in (85) and (84), respectively. Finally, we obtain (81) by substituting $\bar{\Psi}$ and $\bar{\Upsilon}$ in (95) by their respective deterministic equivalents (96) and (100), which completes the proof. ■

V. Sum Rate Analysis Under Limited Feedback

This section analyzes the behavior of the sum rate under the limited-rate feedback link. To obtain tractable expressions, we restrict the subsequent analysis to homogeneous networks ($\mathbf{L}=\mathbf{I}_K$) without transmit correlation ($\mathbf{\Theta}=\mathbf{I}_M$). As previously stated for $\mathbf{L}=\mathbf{I}_K$, we assume $\tau_k=\tau$. In this case and for large (K,M) all users have equal SINR and thus optimizing the SINR is equivalent to optimizing the per-user rate and the sum rate.

The considered performance metric is the sum rate R_{sum} defined in (14). Define R°_{sum} to be

$$R^\circ_{\mathrm{sum}} = \sum_{k=1}^{K}\log_2\left(1+\gamma^\circ_k\right) \quad [\mathrm{bits/s/Hz}], \quad (101)$$

where γ°_k equals $\gamma^\circ_{k,\mathrm{rzf}}$ or $\gamma^\circ_{k,\mathrm{zf}}$ for RZF and ZF precoding, respectively. From $\gamma_k - \gamma^\circ_k \xrightarrow{M\to\infty} 0$, almost surely, we have that $\frac{1}{K}(R_{\mathrm{sum}} - R^\circ_{\mathrm{sum}}) \xrightarrow{M\to\infty} 0$, almost surely. From Remark 1 we infer that, if $\hat{\mathbf{X}}$ is Gaussian we have $R_{\mathrm{sum}} - R^\circ_{\mathrm{sum}} \xrightarrow{M\to\infty} 0$, almost surely.

Subsequently, we will derive the scaling of the distortion τ^2 necessary to approximately maintain a per-user rate gap between perfect CSIT and imperfect CSIT.

A. Optimal Regularized Zero-forcing Precoding

Consider a SNR-independent distortion τ^2. As was observed in [22], [23] for ZF precoding and in [31] for MMSE precoding, the sum rate of ORZF precoding saturates at asymptotically high SNR.

Corollary 3: In the conditions of Corollary 1, the approximate sum rate $R^{\circ,\mathrm{rzf}}_{\mathrm{sum}}$ saturates for asymptotically high SNR at

$$R^{\circ,\mathrm{rzf}}_{\mathrm{lim}} \triangleq \lim_{\rho\to\infty} K\log_2(1+\gamma^\circ_{\mathrm{rzf}}) = K\log_2(1+\bar{\gamma}^\circ_{\mathrm{rzf}}) \quad (102)$$

where $\bar{\gamma}^\circ_{\mathrm{rzf}}$ is given by

$$\bar{\gamma}^\circ_{\mathrm{rzf}} = \begin{cases} -\frac{1}{2} + \sqrt{\frac{1}{\tau^2} - \frac{3}{4}} & \text{if } \beta = 1 \\ \frac{1-\tau^2}{\tau^2}(\beta-1) + \frac{\bar{\chi}}{2} - \frac{1}{2} & \text{if } \beta > 1, \end{cases} \quad (103)$$

where $\bar{\chi}$ is given by (77).

Proof: (102) follows directly from Corollary 1 by taking the limit $\rho\to\infty$. ■

The rate gap per user under ORZF is given in the following theorem.

Theorem 4: Let $\mathbf{\Theta}=\mathbf{I}_M$, $\mathbf{L}=\mathbf{I}_K$ and τ^2 be the distortion per user. Define $\Delta R^{\circ,\mathrm{rzf}}$ to be the difference between the per-user rate of ORZF precoding under perfect CSIT and imperfect CSIT. Then $\Delta R^{\circ,\mathrm{rzf}}$ is given by

$$\Delta R^{\circ,\mathrm{rzf}} = \log_2\left(\frac{1+g(1,\beta)}{1+g(\omega,\beta)}\right) \quad [\mathrm{bits/s/Hz}], \quad (104)$$

where ω is given in (72) and

$$g(x,\beta) = x\rho(\beta-1) + \sqrt{(1-\beta)^2x^2\rho^2 + 2(1+\beta)x\rho + 1}.$$

Proof: With Corollary 1, simply compute $\Delta R^{\circ,\mathrm{rzf}} = R^{\circ,\mathrm{rzf}}(\tau=0) - R^{\circ,\mathrm{rzf}}(\tau)$. ■

Corollary 4: In the conditions of Theorem 4 and with $\beta=1$, the per-user rate loss $\Delta R^{\circ,\mathrm{rzf}}_{\mathrm{sum}}$ takes the form

$$\Delta R^{\circ,\mathrm{rzf}} = \log_2\left(\frac{1+\sqrt{1+4\rho}}{1+\sqrt{1+4\omega\rho}}\right) \quad [\mathrm{bits/s/Hz}]. \quad (105)$$

Following the work in [22], we extend [22, Theorem 3] to ORZF precoding in the following theorem.

Theorem 5: Let $\mathbf{\Theta}=\mathbf{I}_M$ and $\mathbf{L}=\mathbf{I}_K$. To maintain a rate offset no larger than $\log_2 b$ (per user) between ORZF with perfect CSIT and imperfect CSIT, the distortion τ^2 has to scale approximately with

$$\tau^2 = \frac{\phi^{\star\circ}_{\mathrm{rzf}}(\rho,b)}{\rho}, \quad (106)$$

$$\phi^{\star\circ}_{\mathrm{rzf}}(\rho,b) = \frac{\rho\left[(1+\beta)b + w(\beta-1)\right] - \frac{1}{2b}(w^2-b^2)}{(1+\beta)b + w(\beta-1) + \frac{1}{2b}(w^2-b^2)}, \quad (107)$$

$$w(\rho,b) = 1 - b + g(1,\rho). \quad (108)$$

Proof: Set $\Delta R^{\circ,\mathrm{rzf}} = \log_2 b$ and solve for τ^2. ■

Corollary 5: In the conditions of Theorem 5 with $\beta=1$, the distortion τ^2 has to scale approximately with

$$\tau^2 = \frac{1+4\rho-\frac{w^2}{b^2}}{3+\frac{w^2}{b^2}}\frac{1}{\rho}. \quad (109)$$

If the SNR grows to infinity, the term $\phi^{\star\circ}_{\mathrm{rzf}}(\rho,b)$ in (107) converges to the following limits,

$$\lim_{\rho\to\infty}\phi^{\star\circ}_{\mathrm{rzf}}(\rho,b) = \begin{cases} b^2-1 & \text{if } \beta=1 \\ b-1 & \text{if } \beta>1. \end{cases} \quad (110)$$

Details can be found in Appendix VII-A.

For a direct comparison of Theorem 5 to [22, Theorem 3], we set $\tau^2 = 2^{-\frac{B^{\star\circ}_{\mathrm{rzf}}}{M-1}}$, where $B^{\star\circ}_{\mathrm{rzf}}$ is the number of feedback bits per user under ORZF precoding. Thus, (106) takes the form

$$B^{\star\circ}_{\mathrm{rzf}} = (M-1)\log_2\rho - (M-1)\log_2\phi^{\star\circ}_{\mathrm{rzf}}(\rho,b). \quad (111)$$

B. Regularized Zero-forcing Precoding with $\alpha = 1/(\beta\rho)$

Although the RZF-CDU precoder is suboptimal under imperfect CSIT, the results are useful to compare to the work in [22]. In [22, Theorem 3] gives the minimum number of feedback bits necessary to maintain a rate offset of $\log_2 b$ per user for ZF precoding under random vector quantization (RVQ). Moreover, the author claims that [22, Theorem 3] holds also true for RZF-CDU precoding, i.e. the optimal RZF precoder under the assumption of *perfect* CSIT.

For the sake of comparison to [22, Theorem 3], we state the following proposition.

Proposition 3: In the conditions of Theorem 5, under RZF-CDU precoding and $\tau^2 = 2^{-\frac{B^{\circ}_{\rm rzf}}{M-1}}$, the number of feedback bits $B^{\circ}_{\rm rzf}$ has to scale approximately with

$$B^{\circ}_{\rm rzf} = (M-1)\log_2\rho - (M-1)\log_2 \phi^{\circ}_{\rm rzf}(\rho, b), \qquad (112)$$

where for asymptotically high SNR, $\phi^{\circ}_{\rm rzf}(\rho, b)$ converges to the following limits.

$$\lim_{\rho\to\infty} \phi^{\circ}_{\rm rzf}(\rho, b) = \begin{cases} 2(b-1) & \text{if } \beta = 1 \\ b-1 & \text{if } \beta > 1. \end{cases} \qquad (113)$$

The proof is provided in Appendix VII-B.

C. Zero-forcing Precoding

Corollary 6: In the conditions of Corollary 2, the approximate sum rate $R^{\circ,\rm zf}_{\rm sum}$ saturates for asymptotically high SNR at

$$R^{\circ,\rm zf}_{\rm lim} = \lim_{\rho\to\infty} \sum_{k=1}^{K} \log_2\left(1 + \gamma^{\circ}_{k,\rm zf}\right) \qquad (114)$$

$$= K\log_2\left(1 + \frac{1-\tau^2}{\tau^2}(\beta - 1)\right). \qquad (115)$$

Proof: (102) follows directly from Corollary 1 by taking the limit $\rho \to \infty$. ■

Theorem 6: Let $\beta > 1$, $\mathbf{\Theta} = \mathbf{I}_M$, $\mathbf{L} = \mathbf{I}_K$ and τ^2 be the distortion per user. Define $\Delta R^{\circ,\rm zf}$ to be the difference of the per-user rate under ZF precoding of perfect CSIT and imperfect CSIT. Then $\Delta R^{\circ,\rm zf}$ is given by

$$\Delta R^{\circ,\rm zf} = \log_2\left(\frac{1+\rho(\beta-1)}{1 + \frac{(1-\tau^2)(\beta-1)}{\tau^2 + \frac{1}{\rho}}}\right) \quad [\text{bits/s/Hz}]. \qquad (116)$$

Theorem 7: Let $\beta > 1$, $\mathbf{\Theta} = \mathbf{I}_M$ and $\mathbf{L} = \mathbf{I}_K$. To maintain a rate offset no larger than $\log_2 b$ (per user) between ZF with perfect CSIT and imperfect CSIT, the distortion τ^2 has to scale approximately with

$$\tau^2 = \frac{\phi^{\star\circ}_{\rm zf}(\rho, b)}{\rho}, \qquad (117)$$

$$\phi^{\star\circ}_{\rm zf}(\rho, b) = \frac{(b-1)[1+\rho(\beta-1)]}{1-b+(\beta-1)[\rho+b]}. \qquad (118)$$

For asymptotically high SNR, $\phi^{\star\circ}_{\rm zf}(\rho, b)$ in (118) converges to

$$\lim_{\rho\to\infty} \phi^{\star\circ}_{\rm zf}(\rho, b) = b - 1. \qquad (119)$$

Under RVQ we have

$$B^{\star\circ}_{\rm zf} = (M-1)\log_2\rho - (M-1)\log_2\phi^{\star\circ}_{\rm zf}(\rho, b). \qquad (120)$$

D. Discussion

At this point we can draw the following conclusions. For $\beta = 1$ the optimal scaling of the feedback bits $B^{\star\circ}_{\rm rzf}$, $B^{\circ}_{\rm rzf}$ and B for ZF in [22, Theorem 3] are different, even in the high SNR limit. In fact, for RZF-CDU, the upper-bound in [22, Theorem 3] is too pessimistic in the scaling of the feedback bits. From (112) and (113), a more accurate choice is

$$B^{\circ}_{\rm rzf} = (M-1)\log_2\rho - (M-1)\log_2(2(b-1)), \qquad (121)$$

i.e. M–1 bits less than proposed in [22, Theorem 3]. However, notice that (121) is an approximation and not a bound, i.e. a rate gap of $\log_2 b$ bits/s/Hz cannot be guaranteed (as [22, Theorem 3] does) for finite number of users, but is exactly maintained as $K, M \to \infty$. Nevertheless, simulations show that the approximation is accurate for all finite (K, M).

Moreover, for high SNR, to maintain a rate offset of $\log_2 b$ ORZF requires $(M-1)\log_2(b+1)$ bits *less* than ZF precoding and $(M-1)\log_2(\frac{b+1}{2})$ bits *less* than RZF-CDU. Note that the result (110) does not converge to the result for ZF precoding, i.e. $(b-1)$ [22, Theorem 3], since the optimal regularization parameter for ORZF, designed for imperfect CSIT, does not converge to zero, and therefore, the ORZF precoder does not match the ZF precoder for asymptotically high SNR.

In contrast, for $\beta > 1$, we have $B^{\star\circ}_{\rm rzf} = B^{\circ}_{\rm rzf} = B^{\star\circ}_{\rm zf}$ for asymptotically high SNR. Intuitively, the reason is, that for $\beta > 1$ the channel is well conditioned and the RZF and ZF perform very close. Therefore, both schemes are equally sensitive to imperfect CSIT and thus the scaling of τ^2 is the same for high SNR.

Notice that our model comprises a generic distortion of the CSIT. That is, the distortion can be a combination of different additional factors, e.g. channel estimation at the receivers, channel mismatch due to feedback delay or feedback errors. Moreover, we consider i.i.d. block-fading channels, which can be seen as a worst case scenario in terms of feedback overhead. It is possible to exploit channel correlation in time, frequency and space to refine the CSIT or to reduce the amount of feedback.

VI. Applications

This section presents two applications of the theoretical results derived in Section IV.

In what follows, we use the approximated sum rate (101) to compute for ZF, (i) the optimal number of active users selected for transmission and, for ZF and ORZF (ii) the amount of channel training in TDD multi-user systems required to maximize the sum rate. The derived solutions are close approximations to the exact solutions to (i) and (ii) and their accuracy increases as (K, M) grow large. Despite their approximate character, the solutions give valuable insight into the system behavior and can also be utilized as good initialization points for further optimization if (K, M) are small.

A. ZF: Optimal Number of Active Users

For *fixed* $\mathbf{\Theta}$, $\mathbf{L}$, ρ and τ_k^2, we consider the problem of finding the number of users $K^{\star\circ}$ (or equally the system load

$\beta^{\star\circ}=M/K^{\star\circ}$), such that the approximated sum rate (101) is maximized, i.e.

$$\beta^{\star\circ} = \arg\max_{\beta>1} \frac{1}{\beta}\int \log_2\left(1+\gamma^{\circ}_{k,\mathrm{zf}}\right) dF^{\mathbf{L}}(l), \tag{122}$$

where we suppose that the user channel gains l_k are distributed according to some probability distribution function $F^{\mathbf{L}}$. By setting the derivative of (122) w.r.t. β to zero, we obtain the implicit equation

$$\beta \int \frac{\frac{\partial \gamma^{\circ}_{k,\mathrm{zf}}}{\partial\beta} dF^{\mathbf{L}}(l)}{1+\gamma^{\circ}_{k,\mathrm{zf}}} = \int \log_2\left(1+\gamma^{\circ}_{k,\mathrm{zf}}\right) dF^{\mathbf{L}}(l) \tag{123}$$

and $\beta^{\star\circ}$ is a solution to (123).

Proposition 4: Let $\mathbf{\Theta}=\mathbf{I}_M$ and $\mathbf{L}=\mathbf{I}_K$, then $\beta^{\star\circ}$ is given by

$$\beta^{\star\circ} = \left(1-\frac{1}{a}\right)\left(1+\frac{1}{\mathcal{W}(x)}\right), \tag{124}$$

where $\mathcal{W}(x)$ is the single-valued Lambert-W function, defined as the unique solution to $\mathcal{W}(x)=xe^{\mathcal{W}(x)}$, and

$$a = \frac{1-\tau^2}{\tau^2+\frac{1}{\rho}}, \quad x=\frac{a-1}{e}\in[-e^{-1},\infty). \tag{125}$$

The proof of Proposition 4 is given in Appendix VIII. Note that $\lim_{a\to 1}\beta^{\star\circ}=e$. Moreover, only rational values of β are meaningful in practice.

B. Optimal Amount of Channel Training in TDD Multi-user Systems

Consider a time division duplex (TDD) system where uplink and downlink access the *same* channel at different times. Therefore, the transmitter estimates the channel from known pilot signaling of the receivers. The imperfections in the CSIT are caused by (i) channel estimation errors, (ii) imperfect channel reciprocity due to different hardware in the transmitter and receiver and (iii) the duration for which the channel is approximately constant, i.e. the channel coherence time.

In what follows we assume that the channel is perfectly reciprocal and study the impact of (i) and (iii) for $\mathbf{\Theta}=\mathbf{I}_M$ and $\mathbf{L}=\mathbf{I}_K$. The channel is assumed to be constant over $T\in\mathbb{R}^+$ channel uses which are divided into T_d channel uses for data transmission in the downlink and T_t channel uses for channel training in the uplink. During the training phase, each user sends T_t *orthogonal* pilot symbols, (which limits the number of users to $K\leq T_t$), from which the transmitter estimates the channel using MMSE channel estimator. Furthermore, we assume that the SNR ρ_{ul} of the uplink channel is smaller than the SNR ρ_{dl} of the downlink channel, since the receivers usually have less transmit power available. This setup has been considered in [35]–[37].

Since the sum rate is itself depending on T_d *and* T_t, there exists a non-trivial trade off in allocating the channel resources between channel training and data transmission. The approximated normalized sum rate $\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}}=\frac{T_d}{T}R^{\circ,\mathrm{zf}}_{\mathrm{sum}}$ takes the form

$$\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}} = K\left(1-\frac{T_t}{T}\right)\log_2\left(1+\frac{1-\tau^2}{\tau^2+\frac{1}{\rho_{dl}}}(\beta-1)\right). \tag{126}$$

Similarly, for ORZF we have

$$\bar{R}^{\circ,\mathrm{rzf}}_{\mathrm{sum}} = K\left(1-\frac{T_t}{T}\right)\log_2\left(1+\gamma^{\circ}_{\mathrm{rzf}}\right), \tag{127}$$

where $\gamma^{\circ}_{\mathrm{rzf}}$ is given in Corollary 1. The distortion τ^2 in the CSIT is solely caused by an imperfect channel estimation in the uplink and is identical for all entries of $\mathbf{H}$. To acquire CSIT, the users transmit orthogonal pilot symbols over the uplink channel to the transmitter, which receives vector $\mathbf{r}_k$ from user k

$$\mathbf{r}_k = \sqrt{T_t M\rho_{ul}}\mathbf{h}_k + \mathbf{n}_k, \tag{128}$$

where $\mathbf{n}_k\sim\mathcal{CN}(0,\mathbf{I}_K)$ is the noise vector. Subsequently the transmitter performs a MMSE estimation of each channel coefficient. Due to the orthogonality property of the MMSE estimation, the estimate $\hat{h}_{ij}$ of h_{ij} ($i=1,\ldots,K$, $j=1,\ldots,M$), is independent of the estimation error $\tilde{h}_{ij}\sim\mathcal{CN}(0,\sigma_t^2/M)$ and we have [38]

$$h_{ij} = \hat{h}_{ij} + \tilde{h}_{ij}, \tag{129}$$

where the variance $\sigma_t^2/M = E\,\tilde{h}_{ij}\tilde{h}^*_{ij}$ of the estimation error $\tilde{h}_{ij}$ is given by

$$\sigma_t^2 = \frac{1}{1+T_t\rho_{ul}}. \tag{130}$$

Substituting $\tau^2=\sigma_t^2$ into (126) and (127) we obtain

$$\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}} = K\left(1-\frac{T_{t,\mathrm{zf}}}{T}\right)\log_2\left(1+\frac{T_{t,\mathrm{zf}}\rho_{ul}(\beta-1)}{1+T_{t,\mathrm{zf}}\frac{\rho_{ul}}{\rho_{dl}}+\frac{1}{\rho_{dl}}}\right), \tag{131}$$

$$\bar{R}^{\circ,\mathrm{rzf}}_{\mathrm{sum}} = K\left(1-\frac{T_{t,\mathrm{rzf}}}{T}\right)\log_2\left(\frac{1}{2}+\frac{1}{2}w\rho_{dl}(\beta-1)+\frac{d}{2}\right), \tag{132}$$

$$d = \sqrt{(1-\beta)^2 w^2\rho_{dl}^2 + 2w\rho_{dl}(1+\beta)+1}, \tag{133}$$

$$w = \frac{T_{t,\mathrm{rzf}}\rho_{ul}}{1+T_{t,\mathrm{rzf}}\rho_{ul}+\rho_{dl}}. \tag{134}$$

For $\beta>1$ under ZF precoding and $\beta\geq 1$ for RZF precoding, it is easy to verify that the function $\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}}$ and $\bar{R}^{\circ,\mathrm{rzf}}_{\mathrm{sum}}$ are strictly concave in $T_{t,\mathrm{zf}}$ and $T_{t,\mathrm{rzf}}$ in the interval $K\leq T_{t,\mathrm{zf}},T_{t,\mathrm{rzf}}\leq T$, where K is the minimum amount of training due to the orthogonality of the pilot sequences. Therefore, we can apply standard convex optimization algorithms [39] to evaluate

$$T^{\star\circ}_{t,\mathrm{zf}} = \arg\max_{K\leq T_{t,\mathrm{zf}}\leq T}\left\{\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}}\right\}, \tag{135}$$

$$T^{\star\circ}_{t,\mathrm{rzf}} = \arg\max_{K\leq T_{t,\mathrm{rzf}}\leq T}\left\{\bar{R}^{\circ,\mathrm{rzf}}_{\mathrm{sum}}\right\}. \tag{136}$$

The uplink SNR ρ_{ul} and downlink SNR ρ_{dl} are dependent and we consider that ρ_{ul}/ρ_{dl} is constant as both ρ_{ul} and ρ_{dl} grow large. For the limiting cases $\rho_{dl}\to 0$ and $\rho_{dl}\to\infty$, we obtain the following solutions to (135) and (136). Details can be found in Appendix IX.

At asymptotically low SNR, (135) and (136) have the solution (cf. Appendix IX-A)

$$T^{\star\circ}_{t,\mathrm{zf}} = \lim_{\rho_{dl}\to 0}\ \arg\max_{K\leq T_{t,\mathrm{zf}}\leq T}\bar{R}^{\mathrm{zf}}_{\mathrm{sum}} = \frac{T}{2}, \tag{137}$$

$$T^{\star\circ}_{t,\mathrm{rzf}} = \lim_{\rho_{dl}\to 0}\ \arg\max_{K\leq T_{t,\mathrm{rzf}}\leq T}\bar{R}^{\mathrm{rzf}}_{\mathrm{sum}} = \frac{T}{2}. \tag{138}$$

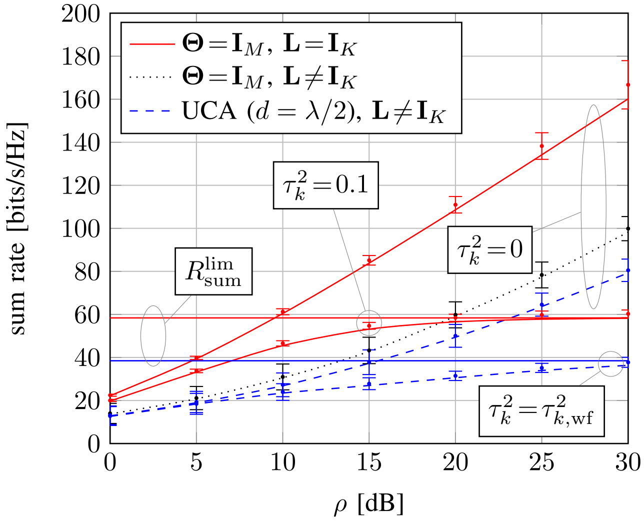

Fig. 1. ORZF, sum rate vs. SNR with $M=32$, $\beta=1$ and $\alpha^{\star\circ}$, simulation results are indicated by circle marks with error bars indicating one standard deviation in each direction.

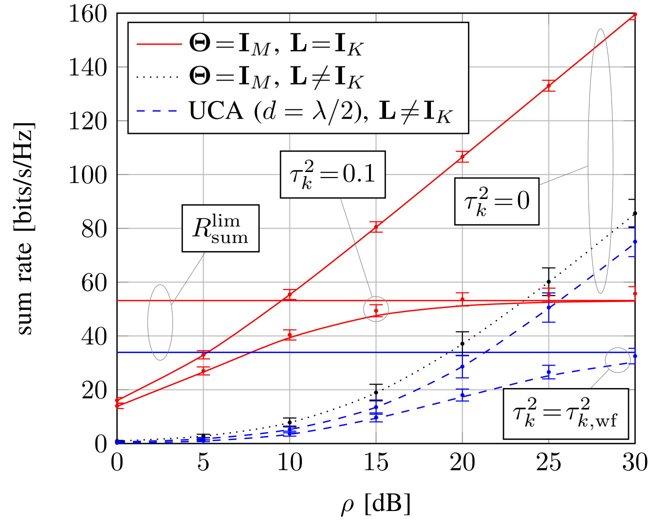

Fig. 2. ZF, sum rate vs. SNR with $M=32$, $\beta=2$ simulation results are indicated by circle marks with error bars indicating one standard deviation in each direction.

That is, the total channel resources T are equally divided into channel training and data transmission.

For asymptotically high SNR the optimal amount of training $T_{t,\text{zf}}^{\star\circ}$ and $T_{t,\text{rzf}}^{\star\circ}$ for ZF precoding and RZF precoding, respectively is given by (cf. Appendix IX-B)

$$T_{t,\text{zf}}^{\star\circ} = \lim_{\rho_{dl}\to\infty} \underset{K\le T_{t,\text{zf}}\le T}{\arg\max}\, \bar{R}_{\text{sum}}^{\text{zf}} = K, \tag{139}$$

$$T_{t,\text{rzf}}^{\star\circ} = \lim_{\rho_{dl}\to\infty} \underset{K\le T_{t,\text{rzf}}\le T}{\arg\max}\, \bar{R}_{\text{sum}}^{\text{rzf}} = K. \tag{140}$$

That is, only the minimal amount of training to obtain a channel estimate is required.

VII. Numerical Results

In this section we compare our theoretical results to Monte-Carlo simulations and assume that the entries of $\mathbf{X}$, $\mathbf{Q}$ are i.i.d. Gaussian distributed.

We begin by introducing the simulation assumptions. Thereafter in Section VII-B, we compare the approximate results of Section IV to Monte-Carlo simulations. The outcome of these experiments clearly reveals the accuracy of the proposed approximations. Afterwards we present simulation results of the ORZF and the two applications discussed in Section VI.

A. Simulation Assumptions

The performance of the Monte-Carlo simulations is averaged over 10,000 independent Rayleigh block-fading channels.

1) Correlation Model: The transmit correlation is assumed to originate from a dense antenna packing at the transmitter. Under a rich scattering environment we use the classical Jakes' model, where the correlation between antenna i and j is depending on their distance d_{ij}, $i,j=1,2,\ldots,M$. Thus, we have [40]

$$(\mathbf{\Theta})_{ij} = \mathrm{J}_0\left(\frac{2\pi d_{ij}}{\lambda}\right), \tag{141}$$

where J_0 is the zero-order Bessel function of the first kind and λ is the signal wavelength. In particular we consider a uniform circular array (UCA) of radius r [41]. To ensure that λ_M grows slower than $O(M)$, we suppose that the distance between adjacent antennas $d=d_{i,i+1}$ is independent of M. Thus, the radius r of the UCA scales with M as $r=d/(2\sin(\pi/M))$. The distance d_{ij} between antennas i and j is given by

$$d_{ij} = 2r\sin\left(\pi\frac{|i-j|}{M}\right). \tag{142}$$

In that case $\mathbf{\Theta}$ is a circulant matrix.

2) Path Loss Model: In order to model the channel gains l_k we consider a circular cell of radius $r_c = 500$ meter and assume that the K users are *uniformly* distributed over the cell area [42]. The path loss is computed according the "Suburban Macro" scenario which is based on the modified COST231 Hata urban propagation model and defined in [43], as

$$l_k = -(31.5 + 35\log_{10} d_k) \quad [\text{dB}], \tag{143}$$

where d_k is the distance of user k to the transmitter. We normalize the l_k s.t. $E\, l_k = 1$ to ensure that the average receive power is identical for all users. The notation $\mathbf{L}\neq\mathbf{I}_K$ indicates that the l_k are distributed according to (143).

B. Deterministic Equivalent of Sum Rate for ORZF and ZF

Figures 1 and 2 compare the sum rate performance of the approximated sum rate to Monte-Carlo simulations for RZF and ZF, respectively. The error bars indicate one standard deviation of the simulation results in each direction. For uncorrelated antennas and equal path loss ($\mathbf{\Theta}=\mathbf{I}_M$, $\mathbf{L}=\mathbf{I}_K$), all τ_k are identical. On the contrary, if $\mathbf{L}\neq\mathbf{I}_K$, we perform a (suboptimal) water-filling algorithm on the per user SNR $\rho_k = l_k\rho/M$ subject to a total of $B=100K$ bits (corresponding approximately to $\tau_k=0.1$, if B is allocated equally among the K users) for RVQ of the user channels. The τ_k^2 obtained by the water-filling algorithm are referred to as $\tau_{k,\text{wf}}^2$.

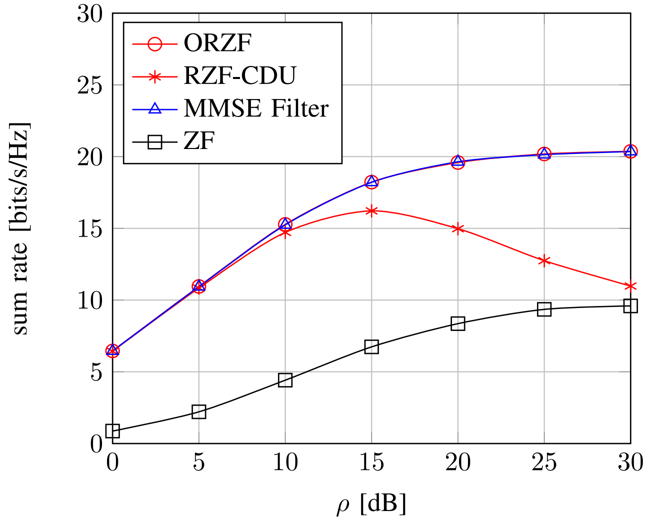

Fig. 3. RZF, ergodic sum rate vs. SNR with $\mathbf{\Theta}=\mathbf{I}_M$, $\mathbf{L}=\mathbf{I}_K$, $M=10$, $\beta=1$, $\tau^2=0.1$.

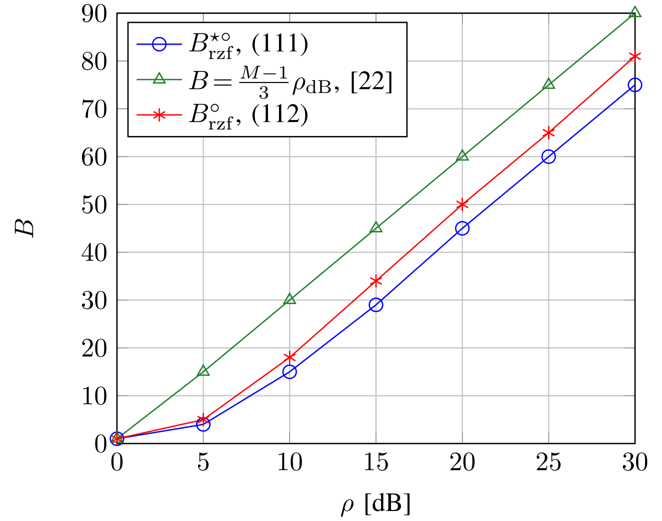

Fig. 5. RZF, B feedback bits per user vs. SNR, with B to maintain a sum rate offset of $K\log_2 b=10$ and $\mathbf{\Theta}=\mathbf{I}_M$, $\mathbf{L}=\mathbf{I}_K$, $M=10$, $\beta=1$.

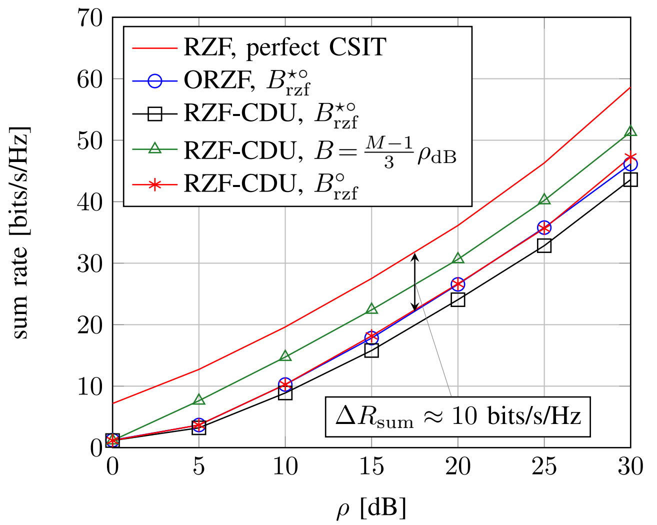

Fig. 4. RZF, ergodic sum rate vs. SNR, with B bits per user to maintain a sum rate offset of $K\log_2 b=10$ and $\mathbf{\Theta}=\mathbf{I}_M$, $\mathbf{L}=\mathbf{I}_K$, $M=10$, $\beta=1$.

We observe that the expressions derived for large (K,M) lie approximately within two standard deviations of the simulation results even for finite (K,M). Therefore, the approximations derived in Section IV are accurate and can be applied to concrete optimization problems for the multi-user downlink channel.

1) RZF under Limited Feedback: Figure 3 shows the ergodic sum rate with error variance $\tau^2=0.1$ for various transmit filters. We consider two RZF filters, ORZF using the sum rate maximizing regularization term $\alpha^{\star\circ}$ in (67) and RZF-CDU with $\alpha=1/(\beta\rho)$, i.e. designed based on perfect CSIT. For comparison we also plot the performance of the MMSE filter proposed in [31], which has an identical structure as ORZF but with $\alpha=\tau^2/\beta+1/(\beta\rho)$, and the ZF filter. We observe that as soon as the error variance τ^2 dominates over the noise power σ^2 the ergodic sum rate of the RZF-CDU filter decreases and approaches ZF precoding for high SNR. We further notice that the ORZF and MMSE filters achieve similar performance. The reason is that, for small values of τ^2, both filters are almost identical and the sum rate is not sensitive to small errors in τ^2 in the filter design [31].

Figure 4 depicts the ergodic sum rate of RZF precoding under RVQ with B feedback bits per user. All values of B are rounded to the next higher integer. To avoid an infinitely high regularization term $\alpha^{\star\circ}$, the minimum number of feedback bits is set to one. We observe that the desired sum rate offset of 10 bits/s/Hz is approximately maintained over the given SNR range when B is chosen as in (111) under ORZF precoding. Given an equal number of feedback bits, it can be seen that the ORZF precoder achieves a significantly higher sum rate compared to the RZF-CDU for medium and high SNR, e.g. about 2.5 bits/s/Hz at 20 dB. Furthermore, to maintain a sum rate offset of M bits/s/Hz, [22] proposed a feedback scaling of $B=\frac{M-1}{3}\rho_{\text{dB}}$. From Figure 4 we observe that this scaling is very pessimistic, since the sum rate offset is about 6 bits/s/Hz. Thus, for RZF-CDU precoding the scaling of the number of feedback bits proposed in (111) proves to be more accurate.

2) ZF: Optimal Number of Active Users: Figure 6 compares the optimal number of active users $K^{\star\circ}$ in (124) to the optimal number of active users $K^\star$ obtained from Monte-Carlo simulations[2], whereas Figure 7 depicts the impact of a suboptimal number of active users on the ergodic sum rate of the system.

From Figure 6 it can be observed, that (i) the approximated results $K^{\star\circ}$ do fit well with the simulation results even for small dimensions, (ii) $(K^\star, K^{\star\circ})$ increase with the SNR (iii), for $\tau^2\neq 0$, $(K^\star, K^{\star\circ})$ saturate at high SNR and (iv) introducing correlation and path loss leads to larger dispersion of $(K^\star, K^{\star\circ})$ over the selected SNR range.

From Figure 7 we observe that, (i) the deterministic equivalent $K^{\star\circ}$ achieve most of the sum rate and (ii) adapting

[2]Note that we do not perform any user scheduling i.e. we do not test all possible combinations of $K\subseteq M$ since that would alter the effective channel distribution.

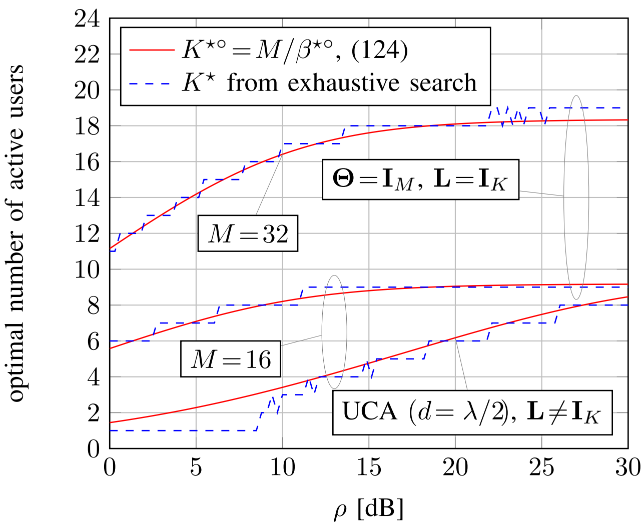

Fig. 6. ZF, sum rate maximizing number of active users vs. SNR with $\tau^2 = 0.1$.

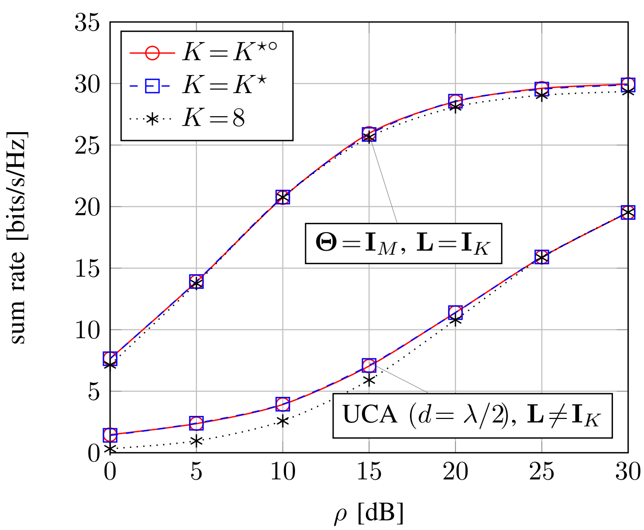

Fig. 7. ZF, ergodic sum rate vs. SNR with $M = 16$ and $\tau^2 = 0.1$.

the number of users is beneficial compared to a fixed K. Moreover, from Figure 6 we identify $K = 8$ as a good choice (for $\tau^2 = 0.1$) and, as expected, the performance is optimal in the medium SNR regime and suboptimal at low and high SNR. The situation changes by adding correlation and path loss. Since $K = 8$ is highly suboptimal for low and medium SNR (cf. Figure 6) we observe a significant loss in sum rate in this regime. Consequently, the number of active users must be adapted to the channel conditions and the approximate result $K^{\star\circ}$ is a good choice to determine the number of active users in the system.

3) Optimal Amount of Channel Training in TDD Multi-user Systems: Figure 8 depicts the optimal relative amount of training $T_t^{\star\circ}/T$ for ZF and ORZF precoding. We observe that $T_t^{\star\circ}/T$ decreases with (i) increasing SNR and (ii), for fixed SNR, with increasing T. That is, for increasing SNR the estimation becomes more accurate and resources for channel training are reallocated to data transmission. Furthermore,

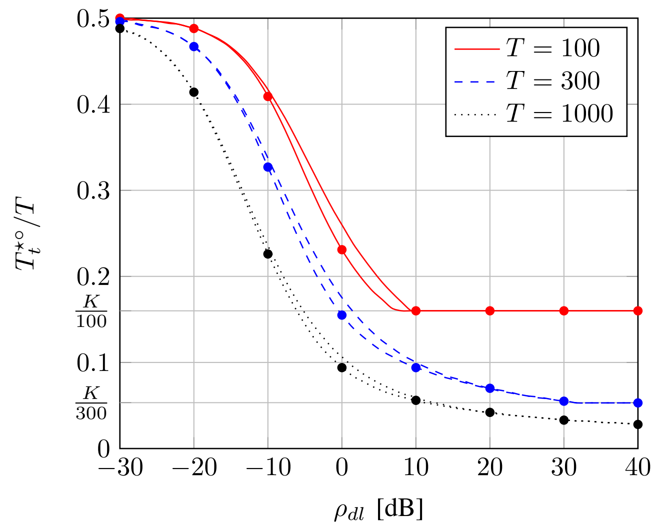

Fig. 8. Optimal amount of training for ZF and ORZF, $T_t^{\star\circ}/T$ vs. downlink SNR with $\mathbf{\Theta} = \mathbf{I}_M$, $\mathbf{L} = \mathbf{I}_K$, $M = 32$, $K = 16$, $\rho_{ul} = 0.1\rho_{dl}$, ORZF is indicated by circle marks.

$T_t^{\star\circ}/T$ saturates at K/T due to the orthogonality constraint on the pilot sequences. Note that $T_t^{\star\circ}/T$ depends on the ratio β and only the saturation level depends on a specific number of users K. The fact that increasing the channel coherence block T reduces the relative amount of training $T_t^{\star\circ}/T$ has also been observed in [44] under a different model for τ^2.

Furthermore, we observe that the optimal amount of training is less for ORZF than for ZF precoding. This is due to the distortion-aware design of the ORZF. Moreover, the relative amount of training $T_t^{\star\circ}/T$ of both ZF and ORZF converges for high and low SNR, to $1/2$ and K, respectively, as predicted by the theoretical analysis.

Figure 9 compares the normalized ergodic sum rates of ZF $\bar{R}_{\text{sum}}^{\text{zf}}$ with optimal training $T_{t,\text{zf}}^{\star\circ}$ for a fixed number of users $K = 16$ and the optimal number of users $K = K^{\star\circ}$ from (124). Since, for a fixed M, $T_{t,\text{zf}}^{\star\circ}$ depends on β and K and $\beta^{\star\circ}$ depends on T_t (through τ) we apply an iterative algorithm to find a joint optimum $(T_{t,\text{zf}}^{\star\circ}, \beta^{\star\circ})$. That is, we fix $T_{t,\text{zf}}$ while optimizing β and vice versa until convergence, which normally occurs after a few iterations. We observe a significant gain in the high SNR regime when K is optimized jointly with T_t. This gain increases for increasing T, since small values of T limit T_t to K (cf. Figure 8) and thus constrain the optimization. On the other hand, for higher values of T (e.g. $T = 1000$) this saturation effect occurs at much higher SNR and T_t is no longer limited by K which increase the gain relative to a fixed K.

VIII. Conclusion

In this paper we derived deterministic equivalents of the SINR of ZF and RZF precoding by applying recent results from large dimensional random matrix theory. These approximations are valid for *any* SNR and shown to be very accurate even for finite dimensions. Therefore, they provide useful tools for many applications, such as the sum rate maximizing number of users in a cell or the optimal ratio

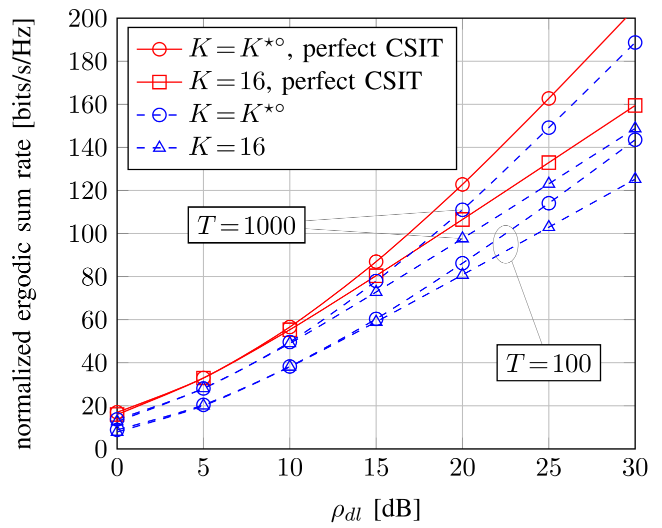

Fig. 9. ZF, approximated sum-rate $\bar{R}^{\circ,\mathrm{zf}}_{\mathrm{sum}}$ vs. forward SNR for $T_{t,\mathrm{zf}} = T^{\star\circ}_{t,\mathrm{zf}}$ with $M = 32$, and $\rho_{ul} = 0.1\rho_{dl}$.

between channel training and data transmission. In particular, we propose a RZF precoder which takes the information about transmit correlation, path loss and CSIT mismatch into account. Furthermore, we find that, given a target user rate, the improved precoder design can significantly decrease the amount of necessary feedback rate.

Appendix I
Proof of Theorem 1

A. Proof of convergence

The proof is an adaptation of the proof provided in [34]. We find a deterministic approximation $m^{\circ}_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ of $m_{\mathbf{B}_N,\mathbf{Q}_N}(z)$ of the form $\mathrm{Tr}\mathbf{D}^{-1}$, such that

$$\mathrm{Tr}\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \mathrm{Tr}\mathbf{D}^{-1} \xrightarrow{N\to\infty} 0, \tag{144}$$

almost surely. Let $\mathbf{T}_N = \mathrm{diag}(\tau_1,\dots,\tau_n)$ and $\mathbf{B}_N = \mathbf{S}_N + \sum_{j=1}\tau_j\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}$ with $\mathbf{y}_j = (1/\sqrt{n})\mathbf{R}_N^{1/2}\mathbf{x}_j$. Now we apply Lemma 8 to compute

$$\begin{aligned}&\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \mathbf{D}^{-1} = \\ &\mathbf{D}^{-1}\left[\mathbf{D} - (\mathbf{B}_N - z\mathbf{I}_N)\mathbf{Q}_N^{-1}\right]\mathbf{Q}_N\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}.\end{aligned} \tag{145}$$

We choose $\mathbf{D} = (\mathbf{S}_N - z\mathbf{I}_N - zp_N\mathbf{R}_N)\,\mathbf{Q}_N^{-1}$ with

$$p_N = -\frac{1}{z\beta n}\sum_{j=1}^{n}\frac{\tau_j}{1+\tau_j e(z)}. \tag{146}$$

Similarly to [34], by taking the trace and dividing by $(1/N)$, equation (145) takes the form

$$\mathrm{Tr}\mathbf{D}^{-1} - m_{\mathbf{B}_N,\mathbf{Q}_N,}(z) = \frac{1}{n}\sum_{j=1}^{n}\tau_j d_j^m = w_M^m \tag{147}$$

$$\text{and}\quad \mathrm{Tr}\mathbf{D}^{-1}\mathbf{R}_N - e(z) = \frac{1}{n}\sum_{j=1}^{n}\tau_j d_j^e = w_M^e, \tag{148}$$

where

$$\begin{aligned}d_j^a = &\frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}_N^{1/2}(\mathbf{B}_{[j]} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{R}_N^{1/2}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{[j]} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\ &- \frac{(1/N)\mathrm{tr}\mathbf{R}_N(\mathbf{B}_{[j]} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+\tau_j e(z)}\end{aligned} \tag{149}$$

with $\mathbf{B}_{[j]} = \mathbf{B}_N - \tau_j\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}$ and $\mathbf{E} = \mathbf{I}_N$ when a is m and $\mathbf{E} = \mathbf{R}_N$ when a is e. The rest of the proof unfolds as in [34], where the authors consider $\mathbf{Q}_N = \mathbf{I}_N$. Since $\mathbf{Q}_N$ is uniformly bounded for all N it does not change the rate of convergence of w_M^m and w_M^e in (147). Thus (144) holds true and $\mathrm{Tr}\mathbf{D}^{-1}$ is a deterministic equivalent of $m_{\mathbf{B}_N,\mathbf{Q}_N}(z)$.

The proof of the existence of $e(z)$ is equivalent to that given in [34]. Furthermore, the uniqueness of $e(z)$ for $\Im(z) > 0$ has also been proved in [34]. Since we evaluate $m_{\mathbf{B}_N,\mathbf{Q}_N}(z)^{\circ}$ for $z < 0$ we prove the uniqueness of $e(z)$ for $z < 0$ in the following subsection.

B. Proof of Uniqueness of $e(z)$

Let $(c, \bar{c})$ and $(e, \bar{e})$ be two solutions of equations (4) and (5), respectively. We prove the uniqueness of $e(z)$ by showing that $e - \bar{e} = 0$. Denote $\mathbf{A} = \mathbf{R}_N^{1/2}(c\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_N^{1/2}$, $\bar{\mathbf{A}} = \mathbf{R}_N^{1/2}(\bar{c}\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_N^{1/2}$, $\mathbf{B} = \mathbf{T}_N^{1/2}(\mathbf{I}_n + e\mathbf{T}_N)^{-1}\mathbf{T}_N^{1/2}$ and $\bar{\mathbf{B}} = \mathbf{T}_N^{1/2}(\mathbf{I}_n + \bar{e}\mathbf{T}_N)^{-1}\mathbf{T}_N^{1/2}$. With Lemma 8, we have

$$c - \bar{c} = -(e - \bar{e})\frac{1}{\beta}\mathrm{Tr}\mathbf{B}\bar{\mathbf{B}}, \tag{150}$$

$$e - \bar{e} = -(c - \bar{c})\mathrm{Tr}\mathbf{A}\bar{\mathbf{A}}. \tag{151}$$

Substituting (150) into (151) we obtain

$$(e - \bar{e})\left(1 - \frac{1}{\beta}\mathrm{Tr}\mathbf{A}\bar{\mathbf{A}}\mathrm{Tr}\mathbf{B}\bar{\mathbf{B}}\right) = 0. \tag{152}$$

In order to show that $e - \bar{e} = 0$, it is sufficient to show that $\eta \triangleq \frac{1}{\beta}\mathrm{Tr}\mathbf{A}\bar{\mathbf{A}}\mathrm{Tr}\mathbf{B}\bar{\mathbf{B}} < 1$. Applying the inequality $|\mathrm{Tr}\mathbf{X}\mathbf{Y}| \le \sqrt{\mathrm{Tr}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathrm{Tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$ to η, we obtain

$$\eta \le \sqrt{\eta_1}\sqrt{\eta_2}, \tag{153}$$

where $\eta_1 = \frac{1}{\beta}\mathrm{Tr}\mathbf{A}\mathbf{A}^{\mathsf{H}}\mathrm{Tr}\mathbf{B}\mathbf{B}^{\mathsf{H}}$ and $\eta_2 = \frac{1}{\beta}\mathrm{Tr}\bar{\mathbf{A}}\bar{\mathbf{A}}^{\mathsf{H}}\mathrm{Tr}\bar{\mathbf{B}}\bar{\mathbf{B}}^{\mathsf{H}}$. It remains to prove that $\eta_1, \eta_2 < 1$. We rewrite (5) and (4) as

$$e = c\mathrm{Tr}\mathbf{A}\mathbf{A}^{\mathsf{H}} + v_1, \tag{154}$$

$$c = e\frac{1}{\beta}\mathrm{Tr}\mathbf{B}\mathbf{B}^{\mathsf{H}} + v_2, \tag{155}$$

where $v_1 = \mathrm{Tr}\mathbf{R}_N^{1/2}(c\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}(\mathbf{S}_N - z\mathbf{I}_N)(c\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_N^{1/2}$ and $v_2 = \frac{1}{\beta}\mathrm{Tr}\mathbf{T}_N(\mathbf{I}_n + e\mathbf{T}_N)^{-2}$. Notice that $v_1, v_2 > 0$ since $z < 0$ and the trace of a Hermitian positive-definite matrix is always positive. Substituting (155) into (154) we obtain

$$e(1 - \eta_1) > 0. \tag{156}$$

For $e > 0$, equation (156) implies that $\eta_1 < 1$. Similarly we have $\eta_2 < 1$ and therefore $\eta < 1$ which completes the proof.

Appendix II
Proof of Proposition 1

The strategy is as follows: First we show the sure convergence of the sequence $\{e_k\}$ for $0 < e_0 \leq -1/z$ to $e(z)$ if $0<|z|<\sqrt{\frac{1}{\beta}\|\mathbf{R}_N\|^2\|\mathbf{T}_N\|^2}$. Then we apply Vitali's convergence theorem [45, p. 169] to extend the sure convergence to all $z<0$ by proving that all e_k are images of Stieltjes transforms at z.

The first step is similar to the procedure in Appendix I-B. Denote $\mathbf{A}_{k-1} = \mathbf{R}_N^{1/2}(c_{k-1}\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_N^{1/2}$, $\mathbf{A}_n = \mathbf{R}_N^{1/2}(c_n\mathbf{R}_N + \mathbf{S}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_N^{1/2}$, $\mathbf{B}_{k-1} = \mathbf{T}_N^{1/2}(\mathbf{I}_n + e_{k-1}\mathbf{T}_N)^{-1}\mathbf{T}_N^{1/2}$ and $\mathbf{B}_{k+1} = \mathbf{T}_N^{1/2}(\mathbf{I}_n + e_k\mathbf{T}_N)^{-1}\mathbf{T}_N^{1/2}$. From (8) and (9) we obtain

$$\frac{e_k - e_{k+1}}{e_{k-1}-e_k} = \frac{1}{\beta}\mathrm{Tr}\mathbf{A}_{k-1}\mathbf{A}_n\mathbf{B}_{k-1}\mathbf{B}_k \triangleq \eta, \tag{157}$$

which can be upper-bounded as

$$\eta \leq \frac{1}{\beta}\|\mathbf{R}_N\|^2\|\mathbf{T}_N\|^2\frac{1}{|z|^2}. \tag{158}$$

If $|z|<\sqrt{\frac{1}{\beta}\|\mathbf{R}_N\|^2\|\mathbf{T}_N\|^2}$ then $\eta<1$ and $\{e_k\}$ is a Cauchy sequence which converges surely.

Suppose e_k is a Stieltjes transforms at z. To prove that e_{k+1} is a Stieltjes transform, we need to verify the following conditions [46, Proposition 2.2]: (i) $\Im[e_{k+1}(z)] > 0$ (ii) $z\Im[e_{k+1}(z)]>0$ and (iii) $\lim_{y\to\infty}[iye_{k+1}(iy)]<\infty$, where $y=\Im[z]$. We rewrite (8) as

$$e_{k+1} = \frac{1}{\beta}\mathrm{Tr}\mathbf{A}_k\mathbf{A}_k^{\mathsf{H}}\mathbf{B}_K\mathbf{B}_K^{\mathsf{H}}e_k + \frac{1}{\beta}\mathrm{Tr}\mathbf{A}_k\mathbf{A}_k^{\mathsf{H}}v_1 + v_2, \tag{159}$$

where $v_1 = \mathrm{Tr}\mathbf{T}_N(\mathbf{I}_n + e_k\mathbf{T}_N)^{-1}(\mathbf{I}_n + e_k^*\mathbf{T}_N)^{-1}$ and $v_2 = \mathrm{Tr}\mathbf{R}_N(c_k\mathbf{R}_N + \mathbf{S} - z\mathbf{I}_N)^{-1}(c_k^*\mathbf{R}_N + \mathbf{S} - z^*\mathbf{I}_N)^{-1}(\mathbf{S}-z^*\mathbf{I}_N)$. It is easy to show that e_{k+1} in (159) verifies all three conditions if $e_k(z)$ is itself a Stieltjes transform.

Since $\{e_k\}$ is a sequence of Stieltjes transforms, it is uniformly bounded and analytic on all compact sets of $\mathbb{C}\backslash\mathbb{R}^+$, in particular on all sets $[-x,-1/x]$, $x>0$. Moreover, $(0,\sqrt{\frac{1}{\beta}\|\mathbf{R}_N\|^2\|\mathbf{T}_N\|^2})$ contains an infinite countable number of points. Therefore, we can apply Vitali's convergence theorem [45, p. 169] which proves that $\{e_k\}$ is surely converging to the unique solution $e(z)$ for all $z<0$.

Appendix III
Proof of Lemma 1

Denote $\mathbf{V} = \left(\mathbf{A} + c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}}\right)^{-1}$. Now $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x}$ can be resolved using Lemma 8

$$\begin{aligned}\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} - \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x} &= \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\left(\mathbf{V}^{-1}-\mathbf{A}\right)\mathbf{A}^{-1}\mathbf{x} = \\ &- \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}(c_0\mathbf{x}\mathbf{x}^{\mathsf{H}} + c_1\mathbf{y}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{x}\mathbf{y}^{\mathsf{H}} + c_2\mathbf{y}\mathbf{x}^{\mathsf{H}})\mathbf{A}^{-1}\mathbf{x}.\end{aligned} \tag{160}$$

Equation (160) can be rewritten as

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x} - \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y}(c_1\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x} + c_2\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x})}{1 + c_0\mathbf{x}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x} + c_2\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}}. \tag{161}$$

Similarly to (160), we apply Lemma 8 to $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y}$. Thus, we obtain an expression involving the terms $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{x}$, $\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{y}$, $\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{A}^{-1}\mathbf{y}$ and $\mathbf{y}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{x}$. To complete the proof we apply Corollary 7 and Corollary 6, with $u=\frac{1}{N}\mathrm{tr}\mathbf{A}^{-1}$ and $u'=\frac{1}{N}\mathrm{tr}\mathbf{U}\mathbf{A}^{-1}$ and we have

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{x} - \frac{u'(1+c_1u)}{(c_0c_1-c_2^2)u^2 + (c_0+c_1)u+1} \overset{N\to\infty}{\longrightarrow} 0, \tag{162}$$

almost surely. Similarly we have

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{V}\mathbf{y} - \frac{-c_2uu'}{(c_0c_1-c_2^2)u^2 + (c_0+c_1)u+1} \overset{N\to\infty}{\longrightarrow} 0, \tag{163}$$

almost surely. Note that as $c_0,c_1,c_2\in\mathbb{R}^+$ and $c_0c_1\geq c_2^2$, equations (162) and (163) hold since then $((c_0c_1-c_2^2)u^2 + (c_0+c_1)u+1)$ is bounded away from zero.

Appendix IV
Important Lemmas

Lemma 4: [47, Lemma 2.2] Let $\mathbf{U}$ be an $N\times N$ invertible matrix and $\mathbf{x}\in\mathbb{C}^N$, $c\in\mathbb{C}$ for which $\mathbf{U}+c\mathbf{x}\mathbf{x}^{\mathsf{H}}$ is invertible. Then

$$\mathbf{x}^{\mathsf{H}}\left(\mathbf{U}+c\mathbf{x}\mathbf{x}^{\mathsf{H}}\right)^{-1} = \frac{\mathbf{x}^{\mathsf{H}}\mathbf{U}^{-1}}{1+c\mathbf{x}^{\mathsf{H}}\mathbf{U}^{-1}\mathbf{x}}. \tag{164}$$

Lemma 5: [48, Lemma 2.7] Let $\mathbf{U}\in\mathbb{C}^{N\times N}$ be a complex matrix with uniformly bounded spectral norm, with respect to N. Let $\mathbf{x}\in\mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and finite 8th order moment. Then,

$$E\left[\mid \mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{x} - \frac{1}{N}\mathrm{tr}\mathbf{U}\mid^4\right] \leq \frac{c}{N^2}, \tag{165}$$

where c is a constant independent of N and $\mathbf{U}$.

Corollary 7: In the conditions of Lemma 5,

$$\mathbf{x}^{\mathsf{H}}\mathbf{U}\mathbf{x} - \frac{1}{N}\mathrm{tr}\mathbf{U} \overset{N\to\infty}{\longrightarrow} 0, \tag{166}$$

almost surely.

Lemma 6: Let $\mathbf{U}$ be as in Lemma 5 and $\mathbf{x},\mathbf{y}\in\mathbb{C}^N$ be mutually independent with standard i.i.d. entries of zero mean, variance $1/N$ and uniformly bounded 4th order moment. Then,

$$\mathbf{y}^{\mathsf{H}}\mathbf{U}\mathbf{x} \overset{N\to\infty}{\longrightarrow} 0, \tag{167}$$

almost surely.

Proof: Remark that $E\,|\mathbf{y}^{\mathsf{H}}\mathbf{U}\mathbf{x}|^4 < c/N^2$ for some constant $c>0$ independent of N. The result then unfolds from the Markov inequality and the Borel-Cantelli Lemma [49]. ■

Lemma 7: [50] Let $\zeta>0$, $\mathbf{X}\in\mathbb{C}^{N\times N}$ Hermitian nonnegative definite, $\tau\in\mathbb{R}$ and $\mathbf{z}\in\mathbb{C}^N$. Then

$$\left|\mathrm{tr}\left((\mathbf{X}+\zeta\mathbf{I}_N)^{-1} - (\mathbf{X}+\tau\mathbf{z}\mathbf{z}^{\mathsf{H}}+\zeta\mathbf{I}_N)^{-1}\right)\right| \leq \frac{1}{\zeta}.$$

Lemma 8: Let $\mathbf{U}$ and $\mathbf{V}$ be two invertible complex matrices of size $N\times N$. Then

$$\mathbf{U}^{-1} - \mathbf{V}^{-1} = -\mathbf{U}^{-1}(\mathbf{U}-\mathbf{V})\mathbf{V}^{-1}. \tag{168}$$

Appendix V
Proof of Lemma 2

The proof unfolds from a direct application of Tonelli's Theorem [49]. Denoting $(X, \mathcal{X}, P_X)$ the probability space that generates the series $\mathbf{x}_1, \mathbf{x}_2, \ldots$, we have that for every $\omega \in A$ (i.e. for every realization $\mathbf{A}_1(\omega), \mathbf{A}_2(\omega), \ldots$ with $\omega \in A$), Corollary 7 holds. From Tonelli's Theorem, the space B of couples $(x, \omega) \in X \times \Omega$ for which Corollary 7 holds, satisfies

$$\int_{X\times\Omega} 1_B(x,\omega) dP_{X\times\Omega}(x,\omega) = \int_\Omega \int_X 1_B(x,\omega) dP_X(x) dP(\omega).$$

If $\omega \in A$, $1_B(x,\omega) = 1$ on a subset of X of probability one. Therefore, the inner integral equals one. Since $P(A) = 1$ the outer integral also equals one, and the result is proved.

Appendix VI
Proof of Lemma 3

The proof unfolds similarly as for Lemma V. For $\omega \in B$, the smallest eigenvalue of $\mathbf{B}_N(\omega)$ is uniformly greater than $\varepsilon \triangleq \varepsilon(\omega)$. Therefore $\mathbf{B}_N(\omega)$ and $\mathbf{B}_N(\omega) + \mathbf{v}\mathbf{v}^\mathsf{H}$ are invertible. Taking $z = -\varepsilon(\omega)/2$, we write

$$\begin{aligned}
&\frac{1}{N}\mathrm{tr}\mathbf{A}_N\mathbf{B}_N^{-1}(\omega) = \\
&\frac{1}{N}\mathrm{tr}\mathbf{A}_N\left(\left[\mathbf{B}_N(\omega) - \frac{\varepsilon}{2}\mathbf{I}_N\right] + \frac{\varepsilon}{2}\mathbf{I}_N\right)^{-1}, \text{ and} \\
&\frac{1}{N}\mathrm{tr}\mathbf{A}_N(\mathbf{B}_N(\omega) + \mathbf{v}\mathbf{v}^\mathsf{H})^{-1} = \\
&\frac{1}{N}\mathrm{tr}\mathbf{A}_N\left(\left[\mathbf{B}_N(\omega) + \mathbf{v}\mathbf{v}^\mathsf{H} - \frac{\varepsilon}{2}\mathbf{I}_N\right] + \frac{\varepsilon}{2}\mathbf{I}_N\right)^{-1}.
\end{aligned}$$

Under these notations, $\mathbf{B}_N(\omega) - \frac{\varepsilon}{2}\mathbf{I}_N$ and $\mathbf{B}_N(\omega) + \mathbf{v}\mathbf{v}^\mathsf{H} - \frac{\varepsilon}{2}\mathbf{I}_N$ are still nonnegative definite for all N. Therefore, the rank-1 perturbation Lemma 7, can be applied for this ω. Then, from Tonelli's Theorem [49], in the space that generates the couples $((\mathbf{x}_1, \mathbf{x}_2, \ldots), (\mathbf{B}_1, \mathbf{B}_2, \ldots))$, the subspace, where the rank-1 perturbation lemma applies has probability one, which completes the proof.

Appendix VII
Details on Limiting Cases in Throughput Analysis

A. Optimal Regularized Zero-forcing Precoding

For $\beta = 1$ observe that $g(1, \rho)$ scales as 4ρ. Thus, for $\rho \to \infty$, (107) converges to $b^2 - 1$.

If $\beta > 1$, the term $g(1, \rho)$ takes the form

$$g(1,\rho) = (\beta - 1)\rho + |1 - \beta|\rho\,(1 + o(1)) \overset{\rho\to\infty}{\longrightarrow} 2\rho(\beta - 1). \quad (169)$$

Therefore, for $\rho \to \infty$, (107) converges to $b - 1$.

B. Regularized Zero-forcing Precoding with $\alpha = 1/(\beta\rho)$

With Theorem 2 for $\alpha = 1/(\beta\rho)$ calculate the rate gap per user $\log_2(\frac{1+\gamma_{k,\mathrm{rzf}}(\tau=0)}{1+\gamma_{k,\mathrm{rzf}}(\tau)})$ for $\tau^2 = 2^{\frac{-B^\circ_{\mathrm{rzf}}}{M-1}}$, equate it to $\log_2 b$ and solve for B°_{rzf}. We obtain (112) with

$$\phi^\circ_{\mathrm{rzf}}(\rho, b) = \rho\frac{\Psi(b - 1 - \gamma^*)[(1 + m^\circ)^2\frac{1}{\rho} + 1] + (m^\circ)^2 b}{\Psi(1 - b + \gamma^*)[(1 + m^\circ)^2 - 1] + (m^\circ)^2 b}, \quad (170)$$

where $\gamma^* \triangleq \gamma^\circ_{k,\mathrm{rzf}}(\tau^2 = 0)$.

For $\beta = 1$, the terms c, m°, Ψ and γ^* scale as $1/\sqrt{\rho}$, $\sqrt{\rho}$, $\sqrt{\rho}/2$ and $\sqrt{\rho}$, respectively. Therefore, $\phi^\circ_{\mathrm{rzf}}(\rho, b) \overset{\rho\to\infty}{\longrightarrow} 2(b - 1)$.

If $\beta > 1$, for large ρ, the term c can be expanded as

$$c \sim \frac{1}{\rho\beta(\beta - 1)} + \frac{1}{4\rho^2\beta(1 - \beta)}. \quad (171)$$

The term m° scales as $\frac{1}{\rho\beta(\beta-1)}$ and Ψ and γ^* scale as $\rho(\beta - 1)$. Note that the SINR γ^* converges to the SINR of ZF precoding (87), for $\tau^2 = 0$ and $\rho \to \infty$. With this approximation we obtain $\phi^\circ_{\mathrm{rzf}}(\rho, b) \overset{\rho\to\infty}{\longrightarrow} b - 1$.

Appendix VIII
Proof of Proposition 4

For $\mathbf{\Theta} = \mathbf{I}_M$ and $\mathbf{L} = \mathbf{I}_K$ from Corollary 2, Equation (81) takes the form

$$\gamma^\circ_{k,\mathrm{zf}} = \frac{1 - \tau^2}{\tau^2 + \frac{1}{\rho}}(\beta - 1). \quad (172)$$

For equation (123) we obtain

$$\frac{a\beta}{1 + a(\beta - 1)} = \log_2(1 + a(\beta - 1)), \quad (173)$$

where $a = \frac{1-\tau^2}{\tau^2 + \frac{1}{\rho}}$. Denoting

$$w(\beta) = \frac{a - 1}{a(\beta - 1) + 1} \quad \text{and} \quad x = \frac{a - 1}{e}, \quad (174)$$

we can rewrite (173) as

$$w(\beta)e^{w(\beta)} = x. \quad (175)$$

Notice that $w(\beta) = \mathcal{W}(x)$, where $\mathcal{W}(x)$ is the Lambert W-function defined for every $z \in \mathbb{C}$ as $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$. Therefore, by solving $w(\beta) = \mathcal{W}(x)$ we have

$$\beta^{\star\circ} = \left(1 - \frac{1}{a}\right)\left(1 + \frac{1}{\mathcal{W}(x)}\right). \quad (176)$$

For $\tau \in [0, 1]$, $\beta > 1$ we have $w \geq -1$ and $x \in [-e^{-1}, \infty)$. In this case the relation $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$ is single-valued and $\mathcal{W}(x)$ is well defined.

Appendix IX
Details on Limiting Cases for Optimal Training

A. Low SNR Regime

For ZF, applying Taylor expansion around $\rho_{dl} = 0$, equation (131) can be written as

$$\bar{R}^{\mathrm{zf}}_{\mathrm{sum}} = \left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)\frac{T_{t,\mathrm{zf}}\rho_{dl}\rho_{ul}(\beta - 1)}{\log 2} + o(1). \quad (177)$$

The terms $\bar{R}^{\mathrm{zf}}_{\mathrm{sum}}$ (131) and $\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)T_{t,\mathrm{zf}}\rho_{dl}\rho_{ul}(\beta - 1)$ are both strictly concave in $T_{t,\mathrm{zf}}$ and $o(1)$ is understood to converge uniformly to zero w.r.t. $T_{t,\mathrm{zf}} \in [K, T]$. Therefore,

$$\begin{aligned}
&\arg\max_{T_{t,\mathrm{zf}}} \bar{R}^{\mathrm{zf}}_{\mathrm{sum}} \\
&- \arg\max_{T_{t,\mathrm{zf}}}\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)\frac{T_{t,\mathrm{zf}}\rho_{dl}\rho_{ul}(\beta - 1)}{\log 2} \overset{\rho_{dl}\to 0}{\longrightarrow} 0. \quad (178)
\end{aligned}$$

The maximum of $(1-T_{t,\mathrm{zf}}/T)\,T_{t,\mathrm{zf}}\rho_{dl}\rho_{ul}(\beta-1)$ is in $T/2$.

For RZF precoding, write (134) as $w = cT_{t,\mathrm{rzf}}\rho_{dl} + o(1)$ and (133) as $d = 1 + cT_{t,\mathrm{rzf}}(1+\beta)\rho_{dl}^2 + o(1)$, where $c = \rho_{ul}/\rho_{dl}$. Then, (132) can be written as

$$\bar{R}_{\mathrm{sum}}^{\circ,\mathrm{rzf}} = \left(1 - \frac{T_{t,\mathrm{rzf}}}{T}\right)\frac{T_{t,\mathrm{rzf}}\beta\rho_{dl}^2}{\log 2} + o(1). \tag{179}$$

With the same arguments as for ZF precoding, maximizing $(1-T_{t,\mathrm{rzf}}/T)\,T_{t,\mathrm{rzf}}\beta\rho_{dl}^2$ yields $T_{t,\mathrm{rzf}}^{\star\circ} = T/2$.

B. High SNR Regime

For ZF and for large ρ_{dl}, equation (131) can be written as

$$\bar{R}_{\mathrm{sum}}^{\mathrm{zf}} = K\left(1 - \frac{T_{t,\mathrm{zf}}}{T}\right)\log_2\rho_{dl}\left[1+o(1)\right] \tag{180}$$

If $T_{t,\mathrm{zf}} > K$, then the ratio

$$\frac{K\left(1-\frac{T_{t,\mathrm{zf}}}{T}\right)\log_2\rho_{dl}\left[1+o(1)\right]}{K\left(1-\frac{K}{T}\right)\log_2\rho_{dl}\left[1+o(1)\right]} = \frac{\left(1-\frac{T_{t,\mathrm{zf}}}{T}\right)\left[1+o(1)\right]}{\left(1-\frac{K}{T}\right)\left[1+o(1)\right]} \tag{181}$$

is asymptotically smaller than one. Therefore, $T_{t,\mathrm{zf}}^{\star\circ} = K$.

Likewise, for RZF, by writing $w = cT_{t,\mathrm{rzf}}/(1+cT_{t,\mathrm{rzf}}) + o(1)$ and $d = (\beta-1)\rho_{dl}\frac{cT_{t,\mathrm{rzf}}}{1+cT_{t,\mathrm{rzf}}} + o(1)$, where $c = \rho_{ul}/\rho_{dl}$, (132) takes the form

$$\bar{R}_{\mathrm{sum}}^{\mathrm{rzf}} = K\left(1 - \frac{T_{t,\mathrm{rzf}}}{T}\right)\log_2\rho_{dl} + o(1). \tag{182}$$

With the same arguments as for ZF, $T_{t,\mathrm{rzf}}^{\star\circ} = K$.

References

[1] G. J. Foschini and M. J. Gans, “On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] E. Telatar, “Capacity of Multi-antenna Gaussian Channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[3] D. Gesbert, M. Kountouris, R. W. Heath, C. B. Chae, and T. Sälzer, “From Single User to Multiuser Communications: Shifting the MIMO Paradigm,” *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, May 2007.

[4] G. Caire and S. Shamai, “On the Achievable Throughput of a Multi-antenna Gaussian Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[5] P. Viswanath and D. N. C. Tse, “Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality,” *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[6] W. Yu and J. M. Cioffi, “Sum Capacity of Gaussian Vector Broadcast Channels,” *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[7] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[8] H. Weingarten, Y. Steinberg, and S. Shamai, “The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[9] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, “A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication–Part I: Channel Inversion and Regularization,” *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[10] T. Yoo and A. Goldsmith, “On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[11] A. Wiesel, Y. C. Eldar, and S. Shamai, “Zero-Forcing Precoding and Generalized Inverses,” *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[12] S. Christensen, R. Agarwal, and J. M. Cioffi, “Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[13] S. Shi, M. Schubert, and H. Boche, “Rate Optimization for Multiuser MIMO Systems With Linear Processing,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[14] B. Hochwald and S. Vishwanath, “Space-Time Multiple Access: Linear Growth in the Sum Rate,” in *Proc. IEEE Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, Oct. 2002, pp. 387–396.

[15] H. Viswanathan and S. Venkatesan, “Asymptotics of Sum Rate for Dirty Paper Coding and Beamforming in Multiple-Antenna Broadcast Channels,” in *Proc. IEEE Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Illinois, Oct. 2003, pp. 1064–1073.

[16] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, “Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems,” in *Proc. IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA’02)*, vol. 1, Prague, Czech Republic, Sep. 2002, pp. 9–13.

[17] A. Wiesel, Y. C. Eldar, and S. Shamai, “Linear Precoding via Conic Optimization for Fixed MIMO Receivers,” *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[18] V. K. Nguyen and J. S. Evans, “Multiuser Transmit Beamforming via Regularized Channel Inversion: A Large System Analysis,” in *Proc. IEEE Global Communications Conference GLOBECOM’08*, New Orleans, LO, Dec. 2008, pp. 1–4.

[19] R. Zakhour and S. V. Hanly, “Base station cooperation on the downlink: Large system analysis,” Jun. 2010. [Online]. Available: http://arxiv.org/abs/1006.3360

[20] R. Couillet, S. Wagner, M. Debbah, and A. Silva, “The Space Frontier: Physical Limits of Multiple Antenna Information Transfer,” in *Proceedings of the 3rd ACM International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS’08)*, no. 84, Athens, Greece, Oct. 2008.

[21] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, “An Overview of Limited Feedback in Wireless Communication Systems,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[22] N. Jindal, “MIMO Broadcast Channels With Finite-Rate Feedback,” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[23] P. Ding, D. J. Love, and M. D. Zoltowski, “Multiple Antenna Broadcast Channels With Shape Feedback and Limited Feedback,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.

[24] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Delft, Netherlands: Now Publishers Inc., 2004.

[25] V. A. Marčenko and L. A. Pastur, “Distribution of Eigenvalues for Some Sets of Random Matrices,” *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.

[26] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, “A New Approach for Mutual Information Analysis of Large Dimensional Multi-Antenna Channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[27] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, “Capacity Scaling in MIMO Wireless Systems Under Correlated Fading,” *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[28] A. M. Tulino, A. Lozano, and S. Verdú, “Impact of Antenna Correlation on the Capacity of Multiantenna Channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[29] H. Shin, M. Z. Win, and M. Chiani, “Asymptotic Statistics of Mutual Information for Doubly Correlated MIMO Channels,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 562–573, Feb. 2008.

[30] F. Kaltenberger, M. Kountouris, D. Gesbert, and R. Knopp, “On the Trade-Off Between Feedback and Capacity in Measured MU-MIMO Channels,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4866–4875, Sep. 2009.

[31] A. D. Dabbagh and D. J. Love, “Multiple Antenna MMSE Based Downlink Precoding with Quantized Feedback or Channel Mismatch,” *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, Nov. 2008.

[32] T. Yoo and A. Goldsmith, “Capacity and Power Allocation for Fading MIMO Channels with Channel Estimation Error,” *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[33] A. A. Hutter, E. de Carvalho, and J. M. Cioffi, “On the Impact of Channel Estimation for Multiple Antenna Diversity reception in Mobile OFDM Systems,” in *Proc. IEEE Conference Record of the Asilomar Conference on Signals, Systems, and Computers*, vol. 2, Pacific Grove, CA, Nov. 2000, pp. 1820–1824.

[34] R. Couillet, M. Debbah, and J. W. Silverstein, “A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-user MIMO Channels,” *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0906.3667v3

[35] T. L. Marzetta, “How Much Training Is Required For Multiuser MIMO?” in *Proc. IEEE Conference Record of the Asilomar Conference on Signals, Systems, and Computers ACSSC'06*, Pacific Grove, CA, Oct. 2006, pp. 359–363.

[36] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, “Linear Precoding for Multi-User Mulitple Antenna TDD Systems,” *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0812.0621

[37] U. Salim and D. Slock, “How much Feedback is Required for TDD Multi-Antenna Broadcast Channels with User Selection?” *EURASIP Journal on Advances in Signal Processing*, 2010, to appear.

[38] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, 1994.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*, 6th ed. New York, USA: Cambridge University Press, 2008.

[40] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ: Wiley-IEEE Press, 1994.

[41] T. S. Pollock, T. D. Abhayapala, and R. A. Kennedy, “Antenna Saturation Effects on Dense Array MIMO Capacity,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, Hong Kong, China, Apr. 2003, pp. 361–364.

[42] J. Salo, G. D. Galdo, J. Salmi, P. Kysti, M. Milojevic, D. Laselva, and C. Schneider, *MATLAB Implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996)*, Jan. 2005. [Online]. Available: http://www.tkk.fi/Units/Radio/scm/

[43] H. Huang, “TR 25.996 Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations,” 3GPP, Tech. Rep., Dec. 2008.

[44] M. Kobayashi, G. Caire, and N. Jindal, “How Much Training and Feedback are Needed in MIMO Broadcast Channels?” in *Proc. IEEE International Symposium on Information Theory (ISIT'08)*, Toronto, Canada, Jul. 2008, pp. 2663–2667.

[45] E. C. Titchmarsh, *The Theory of Functions*, 2nd ed. New York, USA: Oxford University Press, 1939.

[46] W. Hachem, P. Loubaton, and J. Najim, “Deterministic Equivalents for Certain Functionals of Large Random Matrices,” *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, Jun. 2007.

[47] J. W. Silverstein and Z. D. Bai, “On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices,” *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, Aug. 1995.

[48] Z. D. Bai and J. W. Silverstein, “No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices,” *Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.

[49] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.

[50] Z. D. Bai and J. W. Silverstein, “On the Signal-to-Interference Ratio of CDMA Systems in Wireless Communications,” *Annals of Applied Probability*, vol. 17, no. 1, pp. 81–101, Feb. 2007.